REMARKS

A. Status of the Claims

Claims 39-50 were pending at the time of the Action. Claims 40, 41, 43, and 48 have been amended. Claim 39 has been canceled. Claims 40 and 41 have been amended to depend from claim 48. Claim 40 was also amended to correct the spelling of "dipyridodiazepinona" (*see e.g.*, Specification, p. 30, ln. 7). Claim 43 has been amended to maintain the proper antecedent basis with claim 48. Support for the amendment to claim 48 can be found in the specification at, for example, page 10, line 24 to page 11, line 5; and page 13, lines 4-9. No new matter has been added by these amendments.

B. The Information Disclosure Statement

The Action indicates that the information disclosure statement filed on November 26, 2003 did not include copies of the listed publications. Applicant notes, however, that copies of the listed publications are not required according to 37 C.F.R. § 1.98(d). As stated in the information disclosure statement, the present application is a divisional application of Serial No. 09/638,833, filed August 14, 2000. Thus, in accordance with Rule 1.98(d) copies of the listed references were not required as they have been previously submitted to the Patent and Trademark Office in prior application Serial No. 09/638,833.

Applicant notes that if the references have been lost from the Patent and Trademark Office's files, Applicant would gladly have provided another copy of the references had Applicant been notified of the loss. Applicant was not, however, notified by the Patent and Trademark Office that the references were lost. Applicant further notes that it would be improper for any new rejection based on a reference cited in the information disclosure statement to be made final, because the reference could have been obtained from the Patent and Trademark Office's files and considered prior to the mailing of the present Office Action.

C.   The Objection Under 37 C.F.R. § 1.75(c)

The Action objects to claims 39-47 under 37 C.F.R. § 1.75(c) as being of improper dependent form because they refer back to a numerically following claim. The MPEP states that "in situations where a claim refers to a numerically following claim and the dependency is clear, both as presented and as it will be renumbered at issue, *all claims should be examined on the merits and no objection as to form need be made.* In such cases, the examiner will renumber the claims into proper order at the time the application is allowed." MPEP § 608.01(n)(F) (emphasis added). Although claims 39-47 depend, either directly or indirectly, from claim 48, their dependency is clear. Applicants, therefore, respectfully request that the objection be withdrawn and that the claims be renumbered at issue.

D.   The Claims Are Definite

The Action rejects claims 39-50 under 35 U.S.C. § 112, second paragraph, as being indefinite. The Action raises three issues in this regard. First, the Action states that it is unclear whether the RT being inactivated is a purified RT or an RT in a viral particle. Second, the Action states that the preamble of claim 48 fails to set forth an objective. Finally, the Action states that the claims are incomplete for omitting essential positive steps.

The current claims clarify that the RT being inactivated is in a viral particle. In addition, the preamble of current claim 48 specifies that it is a method "of eliciting an immune response." Current claim 48 also includes in the last step the positive recitation "wherein an immune response is elicited in the subject," which refers back to the objective of the method recited in the preamble. In view of the above, the current claims are definite. Applicants, therefore, request the withdrawal of this rejection.

E.   The Claims Are Supported by Adequate Written Description in the Specification

The Action rejects claims 39-50 under 35 U.S.C. § 112, first paragraph, as containing subject matter not reasonably described in the specification. In particular, the Action asserts that the claims encompass large numbers of RT proteins and inactivating compounds, however, the specification only provides working examples with HIV RT and a limited number of photolabeling compounds. Applicants traverse this rejection.

To satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventors had possession of the claimed invention. As set forth in detail below, one skilled in the art would reasonably conclude that the inventor had possession of currently claimed invention based on the description provided for in the specification.

HIV viruses are members of the Retroviridae family. A unique aspect of retrovirus replication is the conversion of a single-stranded RNA from the virus genome into a double-stranded DNA molecule that must integrate into the genome of the host cell prior to the synthesis of viral proteins and nucleic acids (Specification, p. 3, ln. 4-12). Accordingly, all retroviruses possess a reverse transcriptase enzyme, which converts the RNA of their genetic material into DNA. Furthermore, since all reverse transcriptases prime the synthesis of new DNA from tRNA, which is a molecule with abundant secondary structure strongly associated with the enzyme, it is generally accepted that the catalytic unit among reverse transcriptases is phylogenetically conserved (*see e.g.*, Flavell, Retroelements, reverse transcriptase and evolution, *Comp. Biochem. Physiol.* vol 110B,N01 pp3-15, 1995; Xiong and Eickbush, Origin and evolution of retroelements based upon their reverse transcriptase sequences, *The EMBO Journal* vol.9 no10 pp3353-3362, 1990; Boeke, The unusual phylogenetic distruibution of retrotransposons: A hypothesis, *Genome Res.* 2003 13:1975-1983; Nakamura *et al.*, Telomerase

catalytic subunit homologs from fission yeast and human, *Science* vol.277, August 15 1997;

Springer *et al.*, Phylogenetic relationships of reverse transcriptase and Rnase H sequences and

aspects of genome structure in the gypsy group of retrotransposons, *Mol. Biol. Evol.* 10(6):1370-

1379, 1993; Lingner *et al.*, Reverse trasncriptase motifs in the catalytic subunit of telomerase,

Science 276:561(1997); Valverde-Garduno *et al.*, Functional analysis of HIV-1 reverse

transcriptase motif C: site directed-mutagenesis and metal cation interaction, *J. Mol. Evol.* 1998

Jul; 47(1):73-80; Seifarth *et al.*, Rapid identification of all known retroviral reverse transcriptase

sequences with a novel versatile detection assay, *AIDS Research and Human Retroviruses*, vol.

16 Number 8, pp 721-729, 2000).

Since retroviruses cannot integrate into the genetic machinery of the host cell without

reverse transcription, the inhibition of reverse transcriptase will have as a universal consequence

the inability of any retrovirus to integrate within the genetic machinery of a suitable host cell.

Thus, regardless of the type of retrovirus, the inactivation of reverse transcriptase as described in

the present specification would be understood by a person of ordinary skill in the art to be

applicable to any retrovirus. Furthermore, the present specification specifically states that "the

methodology of the present invention is applicable to any retrovirus which may be associated

with any animal or human disease as a method for development of effective immunogens and

preventative vaccines. Thus, the present invention has a broader applicability than the

exemplified HIV vaccine." (Specification, pg. 16, ln. 20-23).

As described in the present specification, reverse transcriptase may be inactivated by

binding the reverse transcriptase with one or more azido-labeled compounds and then irradiating

it (*see e.g.,* p. 12, ln. 8-9). The written description requirement for a claimed genus may be

satisfied through sufficient description of a **representative number** of species. Furthermore, it

is not necessary that every permutation within a generally operable invention be effective for

Applicant to obtain a generic claim, provided that the effect is sufficiently demonstrated to characterize a generic invention. *Capon v. Eshhar*, 418 F.3d 1349, 1359 (Fed. Cir. 2005). Numerous compounds that bind to reverse transcriptase are known in the art (*see e.g.*, Specification, p. 3, ln. 23, to p. 4, ln. 11). Many of these compounds are nucleoside analogs, such as AZT, which inhibit the reverse transcriptase by competing with the naturally occurring deoxynucleotides needed to synthesize the viral DNA. There are also non-nucleoside inhibitors (NNIs) of reverse transcriptase. Examples of NNIs provided in the present specification include nevirapine and its analogues, the pyridinones, the pyridobenzo- and dipyridodiazepinones, the quinoxalines, and the carboxanilides (Specification, p. 3, ln. 27, to p. 4, ln. 1; p. 21, ln. 21-26). Additional examples of compounds that bind reverse transcriptase include UC781, 9-AN, UC38, UC84, UC10, UC82, UC040, HBY 097, calanolide A, and U-88204E (*see e.g.*, Specification, pg. 12, ln. 14-18).

As described in the present specification, the above-mentioned compounds may be converted to azido photoaffinity labels and utilized for the inactivation of reverse transcriptase (Specification, p. 21, ln. 21-28). Thus, for example, azido-labeled compounds, such as azido-UC781, and the azido derivatives of 9-AN, UC38, UC84, UC10, UC82, UC040, HBY 097, calanolide A, and U-88204E, may be used to inactivate reverse transcriptase according to the methods described in the present specification. Thus, the present specification satisfies the written description requirement by describing a representative number of azido-labeled compounds that bind reverse transcriptase.

In addition, examples of identifying characteristics of biomolecules listed in the Action (*see* p. 5) include chemical structure and binding affinity. The azido-labeled compounds have, at least, the common structural feature of the azido group and the ability to bind reverse

transcriptase. Thus, the present specification also satisfies the written description requirement by describing the relevant identifying characteristics of the azido-labeled compounds.

In view of the above, the present specification describes the claimed invention in sufficient detail that one of ordinary skill in the art can reasonably conclude that Applicant had possession of the claimed invention at the time of filing.

F.   Conclusion

Applicant believes that this is a full and complete response to the Office Action mailed October 5, 2006. Should the Examiner have any questions, comments, or suggestions relating to this case, the Examiner is invited to contact the undersigned Applicants' representative at (512) 536-5654.

Respectfully submitted,

Travis M. Wohlers
Reg. No. 57,423
Attorney for Applicants

FULBRIGHT & JAWORSKI L.L.P.
600 Congress Avenue, Suite 2400
Austin, Texas 78701
(512) 536-5654
(512) 536-4598 (facsimile)

Date: March 5, 2007

INVITED REVIEW

Retroelements, reverse transcriptase and evolution

Andrew J. Flavell

Department of Biochemistry, Medical Sciences Institute, The University, Dundee DD1 4HN, U.K.

Retroelements are genetic elements that can exist as DNA or RNA or DNA/RNA duplexes. Although retroviruses are the best known retroelements, there are many other types, including close relatives of retroviruses like LTR retrotransposons, more distant relatives like non-LTR retrotransposons, caulimoviruses and hepadnaviruses and elements with virtually no similarity, like retrons. Virtually all retroelements are 'selfish DNAs' with no involvement with the normal development or maintenance of their host cells, the only known exception being telomereres/telomerases which maintain the ends of chromosomes. Virtually all retroelements use tRNA, or RNA with strong secondary structure, to initiate their reverse transcription. The coincidence between the use of tRNA, a molecule central to the conversion of RNA to protein, with reverse transcriptase, an enzyme which is crucial for the conversion of RNA to DNA is striking, because RNA probably preceded DNA and protein in evolution. It seems plausible that retroelements were present at the genesis of living systems.

Introduction

This review is intended to give the reader an overview of the many apparently diverse manifestations of genetic elements containing reverse transcriptase. What emerges from a closer look at these elements is the striking similarity between many of them, suggesting that many, and perhaps all these elements share a common evolutionary origin.

The Origin of Reverse Transcriptase

RNA has an intrinsic catalytic ability to make and break its own phosphodiester backbone. We, therefore, believe that RNA was probably the first self-replicating entity and evolution first worked on it before DNA and protein were brought into the picture (Cech and Bass, 1986; Darnell and Doolittle, 1986).

Although RNA was probably the first genetic material, it is poorly suited to that role because it is chemically labile. DNA is much more inert and better suited to carrying genetic information between generations. RNA can be converted to DNA by reverse transcriptase, an enzyme which is related in sequence to RNA replicases

Table 1. Types of retroelement

Telomeres/telomerase
Group II introns
Retrons
Fungal mitochondrial plasmids
Non-LTR retrotransposons
LTR retrotransposons
Retroviruses
Caulimoviruses and hepadnaviruses

which exist in RNA viruses (Xiong and Eickbush, 1990). We believe that reverse transcriptase evolved from an RNA replicase and this event was central to the development of DNA as the main form of genetic material in living organisms.

Once reverse transcriptase had accomplished this feat, it largely bowed out of the main story of evolution, leaving DNA as the genetic store and RNA as either the message for production of proteins (messenger RNA) or part of the machinery of RNA splicing, polyadenylation, etc, catalysed by ribonucleoprotein complexes, and translation, catalysed by transfer RNAs and ribosomal RNAs. But it did not disappear entirely and has been found still in a wide variety of guises. These 'retroelements' are listed in Table 1 and I will first review them in ascending order of sophistication, concentrating mainly on what we know of their replication cycles, before discussing the evolutionary implications of these data.

Telomeres and Telomerases

Telomeres are the extreme ends of linear chromosomes. Linear chromosomes are ubiquitous (as far as we know) in nuclear eukaryotes genomes and they face the same dilemma found by all linear DNAs, namely, how to resist agents which lead to shortening, such as attack by DNA exonucleases or the inherent inability of DNA polymerase to copy its template to its extreme 5' end. The basic solution to the problem for the majority of eukaryotes is to locate a simple sequence at the telomere which counteracts the reduction in size by replicating extra copies of itself (Blackburn, 1992; Schippen, 1993). The exact sequences of the repeats within telomeres are species-specific, but a typical example is that of *Tetrahymena thermophila*, in which the sequence GGGGGTT is repeated many times.

The template for these extra copies is an RNA molecule which is bound tightly to a specialised type of reverse transcriptase called telomerase. In this case, the template RNA carries a homologous sequence (5'AACCCCAA3') which is used to generate a new DNA strand (Fig. 1).

There is still very little known about the telomerase enzyme which carries out this function, because it is difficult to work with. An exciting possibility is that the RNA template itself is directly involved in the catalytic process, making it another RNA enzyme (ribozyme). However, this hypothesis remains untested and no ribozyme to date has been shown to catalyse reactions on DNA. Perhaps RNA never learned this feat and DNA has only ever been 'genetically manipulated' by proteins.

Group II Introns

Two revolutionarily ancient types of intron survive in the organelles or plasmids of lower eukaryotes. These group I and II introns have the ability to splice themselves out of their precursor mRNAs, without the help of proteins (Cech and Bass, 1986). The two groups are classified by their characteristic sequence motifs which themselves define group-specific conserved secondary structures in the intron RNA. Several of these introns encode 'maturase' genes which aid the splicing process in the cell, although the polypeptides encoded by these genes are not essential for self-splicing in the test tube (Carignani *et al.*, 1983). A variety of polypeptides are used, these are derivatives of enzymes concerned with RNA metabolism (Lambowitz and Perlman, 1990).

Polypeptides resembling aminoacyl tRNA synthetases are common and reverse transcriptase-like proteins are also found. These reverse transcriptase-like proteins have not yet been shown to have enzymic activity but another strange property of group II introns suggests that this may be the case. Some group II introns occasionally transpose to new chromosomal locations (Mueller *et al.*, 1993; Sellem *et al.*, 1993). These new locations are sometimes non-homologous to pre-existing insertion sites for the introns, suggesting that this is true transposition and not a phenomenon related to homologous recombination.

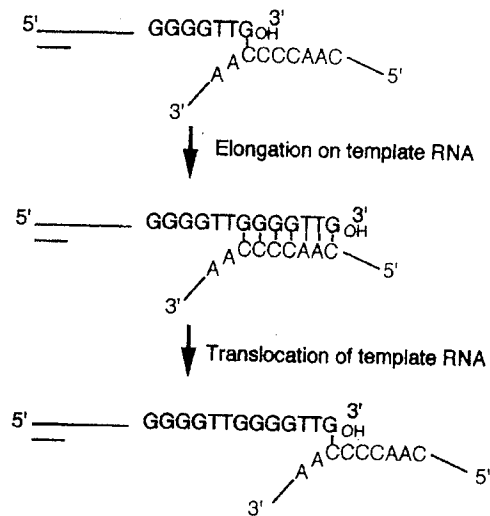Fig. 1. Mechanism for elongation of telomere repeats by telomerase. DNA is shown in black and RNA in red.
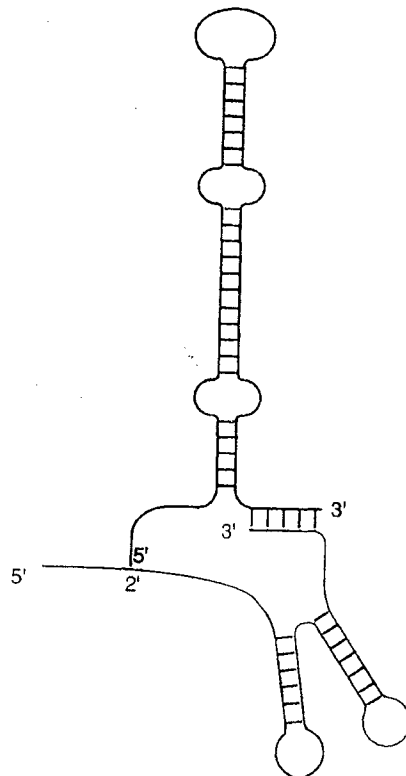

Fig. 2. msDNA. Base pairs involved in secondary structure are shown. DNA is shown in black and RNA in red.
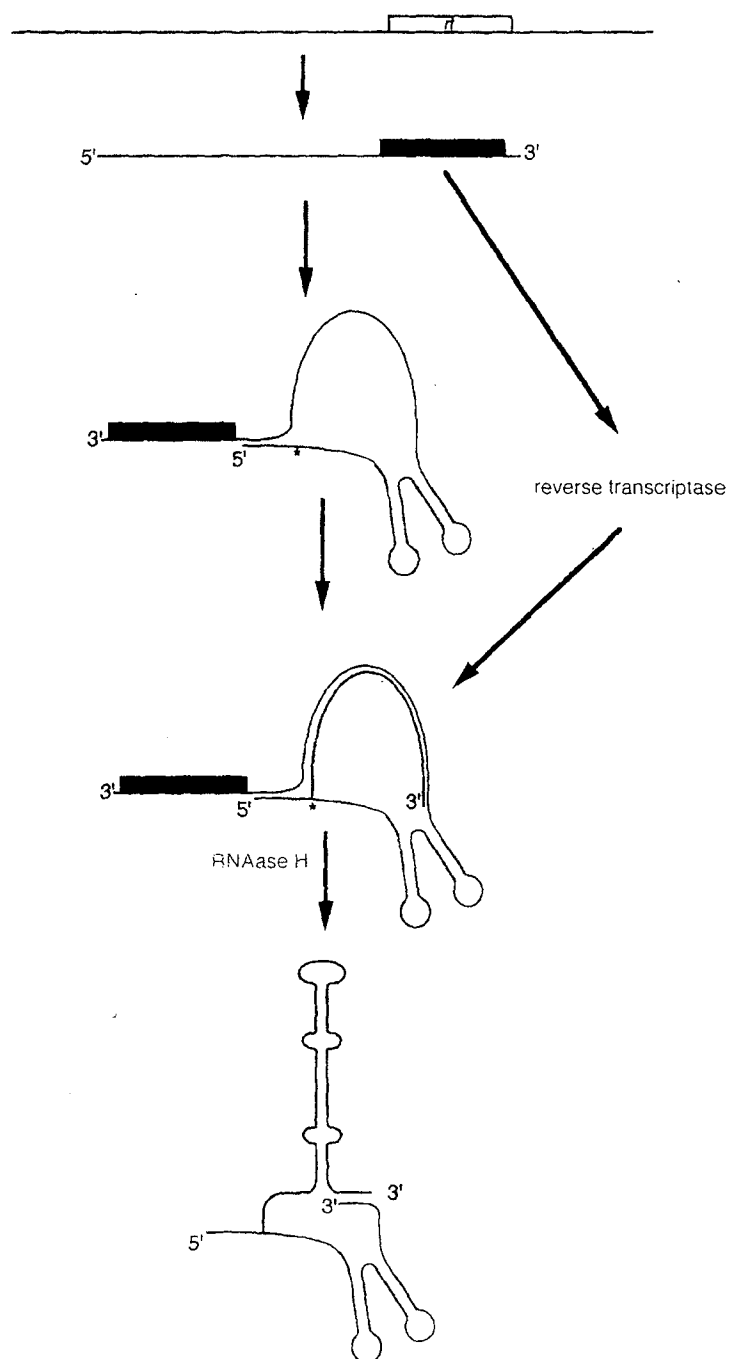
Fig. 3. Biosynthesis of msDNA from its retron template. DNA is shown in black and RNA in red. Initiation of reverse transcription is from the 2′ OH group of a G base (indicated by *).

While other models are possible, it is feasible that the transposition occurs in the following way. First, an RNA copy of the intron becomes inserted into another RNA by a reversal of the normal splicing mechanism (such reverse-splicing has already been demonstrated in group II introns; Augustin *et al.*, 1990). Then the novel intron-containing RNA is reverse transcribed into DNA which then either recombines into the chromosomal DNA directly or by gene conversion.

Retrons

The bizarre RNA–DNA chimaera shown in Fig. 2 is found in many bacteria (Inouye and Inouye, 1993). This small extrachromosomal molecule, called msDNA, is synthesized in bacterial strains containing a genetic element called a retron (Fig. 3). A retron minimally consists of the DNA template for msDNA synthesis, plus a reverse transcriptase gene. A single promoter transcribes the entire genetic element,

Fig. 4. The Mauriceville mitochondrial plasmid of *Neurospora*. (a) Secondary structure of the RNA transcript. RNA is shown by thin lines and DNA by thick lines. Bases conserved with tRNAs are shown by small black circles. (b) *In vitro* initiation of reverse transcription by elongation primed from the 3′ end of the RNA. (c) *In vitro* initiation of reverse transcription by a non-specific DNA oligonucleotide (shown by a striped line). (d) *In vitro* initiation of reverse transcription by elongation primed by G mononucleoside or mononucleotide.

Fig. 5. Gene structural features of retrotransposons and retroviruses. See text for functions of genes.

producing a long transcript from which reverse transcriptase is synthesized. The transcript can adopt a secondary structure, because it contains several regions of self-complementarity. This folded RNA then acts as a template for reverse transcription by the retron-encoded enzyme. This synthesis seems to be primed from a 2' OH group (starred in Fig. 3), unlike all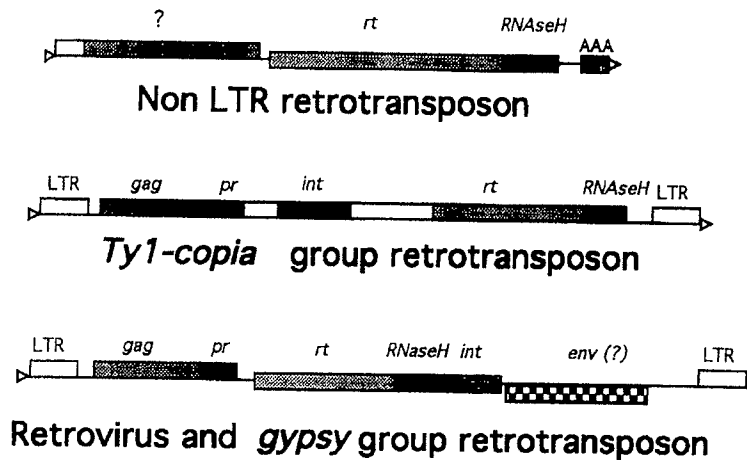 other known reverse transcriptases, RNA polymerases and DNA polymerases, which prime from 3' OH groups. The reverse transcription becomes stalled at one of the RNA hairpin loops. Degradation of most of the RNA part of the resulting heteroduplex leads to the mature msDNA molecule.

What is the point of this strange phenomenon? Only a proportion of the members of a retron-containing bacterial species actually contain retrons. We are therefore confident that this genetic element is a kind of 'selfish DNA' (Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Sapienza and Doolittle, 1981) which is non-essential for the host and whose prime function is its own propagation. It seems unlikely that msDNA is an intermediate in an extrachromosomal replication cycle of retrons, because large parts of the complete retron are missing from it (including the reverse transcriptase gene) but it may be the abortive descendant of such an intermediate. We shall see below that most retroelements use reverse transcription to replicate themselves via extrachromosomal intermediates.

Fungal Mitochondrial Plasmids

Certain fungi (notably some *Neurospora* species) sometimes contain small circular double stranded plasmids in their mitochondria. The best characterized of these (the Mauriceville and Varkud plasmids) have been shown to contain reverse transcriptase genes and to be replicated by a transcription-reverse transcription cycle (Fig. 4). Transcription of the plasmid yields an RNA of exactly the same size as the plasmid, with an intriguing secondary structure at the 3' end which is reminiscent of tRNAs and the 3' ends of RNA viruses of plants and bacteria; indeed, several bases conserved among tRNAs (shown by small black circles in the figure) are found in the plasmid RNA (Fig. 4a). *In vitro* studies using the purified mitochondrial enzyme has shown that reverse transcription can be initiated in three different ways (Wang and Lambowitz, 1993). The first involves elongation from the 3' end of the RNA, perhaps via the secondary structure shown in Fig. 4b. The second way is by elongation of a short, non-specific DNA oligonucleotide which is bound to the reverse transcriptase (Fig. 4c). This time DNA synthesis begins at the penultimate base of the RNA. The final way (Fig. 4d) is identical to the second, except that no exogenous DNA primer is required, just a G base. dGMP, dGDP and dGTP all function in this *in vitro* reaction. It is unclear which, if any of these methods of reverse transcription initiation

predominates *in vivo*. At present, we do not know how the full DNA–RNA hybrid which is synthesized as the first step in the reverse transcription process, is eventually converted into the circular double stranded plasmid.

Non LTR Retrotransposons

Retrotransposons are transposable genetic elements which use reverse transcription during their movement to new chromosomal locations. Non-LTR retrotransposons are the simplest type of retrotransposon. They are found in the genomes of the majority of eukaryotes (including mammals, where they are called LINE, L1 or Kpn elements; Hutchinson *et al.*, 1989; Martin, 1991). They contain two discernible genes, one of which is a reverse transcriptase/RNAse H gene (*rt RNAseH*; Fig. 5). A single transcript, initiating at the exact 5' boundary of the retroelement, encodes all the genetic information of the element (Mizrokhi *et al.*, 1988). The details of the transposition cycle are still unclear but a particle, comprised of proteins encoded by the other gene in the element (marked by ? in the figure) and containing the retrotransposon's RNA and reverse transcriptase, has been identified and is perhaps involved in the reverse transcription and integration process, just as the analogous proteinaceous particles are implicated in the transposition cycles of LTR retrotransposons and retroviruses (see below; Deragon *et al.*, 1990; Martin and Branciforte, 1993). The exact reverse transcription mechanism for non-LTR retrotransposons is still unclear, but is believed that the reverse transcript becomes inserted at random nicks in the chromosome (Finnegan, 1989).

LTR Retrotransposons and Retroviruses

These two types of retroelement are so similar that I will consider them together. The gene structures of retrotransposons and the DNA forms of retroviruses (called proviruses) are shown in Fig. 5. There are two main groups of LTR retrotransposons, the *Tyl-copia* group and *gypsy* group, named after prototype elements in yeast

and *Drosophila*, respectively (Bingham and Zachar, 1989). The former is structurally the more simple group. Both groups are found in fungi, plants and insects. There has been some debate about the existence of retrotransposons in vertebrates. Early claims were probably no more than defective retroviruses (Hodgson *et al.*, 1990) but recent comprehensive PCR-based searches have identified *Tyl-copia* group retrotransposons in fish, amphibia and reptiles (though not mammals and birds as yet; Flavell and Smith, 1992; Flavell *et al.*, 1994).

LTR retrotransposons are known to use intracellular ribonucleoprotein particles as intermediates in their transposition cycles (Shiba and Saigo, 1983; Garfinkel *et al.*, 1985). The protein structural components of the particles are encoded by the *gag* genes (see Fig. 5). The particles also contain virtually full length transcripts of the transposons and several enzymes, all of which are retrotransposon-encoded. A protease (encoded by the *pr* gene) is involved in cleaving the precursor polyprotein into the mature proteins, a reverse transcriptase first copies the transposon RNA to form an RNA–DNA duplex. For all LTR retrotransposons and retroviruses, initiation of the reverse transcription is from a 3' end of tRNA physically bound to the template and the enzyme (Bingham and Xachar, 1989; Varmus and Brown, 1989). The next step is the degradadion of the RNA in the duplex by ribonuclease H to enable synthesis of the second DNA strand by the reverse transcriptase. Finally, an integrase (encoded by the *int* gene) catalyses the insertion of the double-stranded DNA copy into the chromosome.

Retroviruses use just the same genes to achieve the same result as LTR retrotransposons (Varmus and Brown, 1989). The only significant functional difference between the two is the retroviruses proven ability to leave the cell as a virus particle. Entry of virus particles into a new cell requires an envelope glycoprotein, encoded by the *env* gene (Fig. 5), which is embedded in the plasma membrane envelope acquired by the virus when it buds out through the cell membrane. In fact, *gypsy* group retrotransposons are believed by some to actually be retroviruses. They possess an extra

gene of unknown function in the same location as *env* and there is some evidence to suggest that the *gypsy* transposon of *Drosophila* can form an infectious particle (Kim *et al.*, 1994).

Hepadnaviruses and Caulimoviruses (Pararetroviruses)

Two other virus groups use reverse transcription in their life cycles, though neither normally integrate their DNAs into the host chromosomes. Hepadnaviruses (hepatitis B-like viruses) infect vertebrates. They contain a small circular DNA molecule which is partially single-stranded (Fig. 6; Tiollais *et al.*, 1981). This DNA encodes genes specifying a reverse transcriptase, capsid structural components and viral surface proteins. Upon infection, the virion DNA is filled in by the encapsidated reverse transcriptase to form a closed circular double stranded DNA. Transcription of this template yields an RNA which is translated into the virus-encoded proteins (Summers and Mason, 1982). Initiation of reverse transcription is primed by a protein bound to the virion RNA, unlike all other retroelements which use RNA primers (Gerlich and Robinson, 1980). Reverse transcription commences in the particle but remains incomplete, forming the partially single-stranded virion DNA.

Caulimoviruses are plant viruses which share with hepadnaviruses the properties of encapsidation of an incompletely reverse transcribed DNA (Bonneville *et al.*, 1988). In this case, the capsid nucleic acid is largely double-stranded with a few nicks (Fig. 6). The basic steps of transcription from an extrachromosomal closed circular DNA template into RNA which is translated into reverse transcriptase and virus particle components are shared with hepadaviruses. Priming of caulimovirus reverse transcription uses a tRNA, just as LTR retrotransposons and retroviruses do.

The Evolution of Retroelements

From the above survey, it is evident that there is a wide variety of genetic elements which use reverse transcriptase for their propagation in a broad spectrum of living organisms from bacteria to man. I said at the outset that we believe the process of reverse transcription to be evolutionarily ancient. Can we assemble all the known manifestations of reverse transcription and the genetic elements involved with this process into an evolutionary tree? In some cases, this is quite easy (Fig. 7). Retroviruses and LTR retrotransposons are obviously related and the more complex gene structure of the former, plus phylogenetic comparisons of the DNA sequences of these elements (Temin, 1980; Xiong and Eickbush, 1990) suggests strongly that LTR retrotransposons were the ancestors of retroviruses (Fig. 7). Judged by sequence homology and structural similarity, the most likely immediate progeniter of retroviruses was a *gypsy* group LTR retrotransposon. *Ty*1-*copia* group retrotransposons, with their simpler gene structure, probably arose before the *gypsy* group, though whether they were the direct ancestor is unclear.

The hepadnaviruses and caulimoviruses are more difficult to fit into this picture. Phylogenetic analysis suggests that caulimoviruses evolved from *gypsy* group LTR retrotransposons (Doolittle *et al.*, 1989; Xiong and Eickbush, 1990), but hepadnaviruses are highly diverged from both groups. Temin has proposed that both viruses evolved from retroviruses by loss of the ability of the extrachromosomal DNA to integrate into the host chromosome (Temin, 1989). Xiong and Eickbush suggest that hepadnaviruses arose from a recombination event between a pre-existing RNA virus and a primitive retrotransposon while caulimoviruses derived in a similar manner from *gypsy* group retrotransposons. Both models are plausible but the latter is more likely, at least in the case of hepadnaviruses, which have a priming mechanism for the initiation of replication which differs from retroviruses and resembles some RNA viruses, such as poliovirus.

What about the more primitive retroelements? Non-LTR retrotransposons may be the ancestors of LTR retrotransposons, because of their simpler construction, less sophisticated transposition mechanism and ubiquity in the eukaryotes (Doolittle *et al.*, 1989; Xiong and Eickbush, 1990). The evolutionary status of the other elements

HEPADNAVIRUSES   RETROVIRUSES   CAULIMOVIRUSES

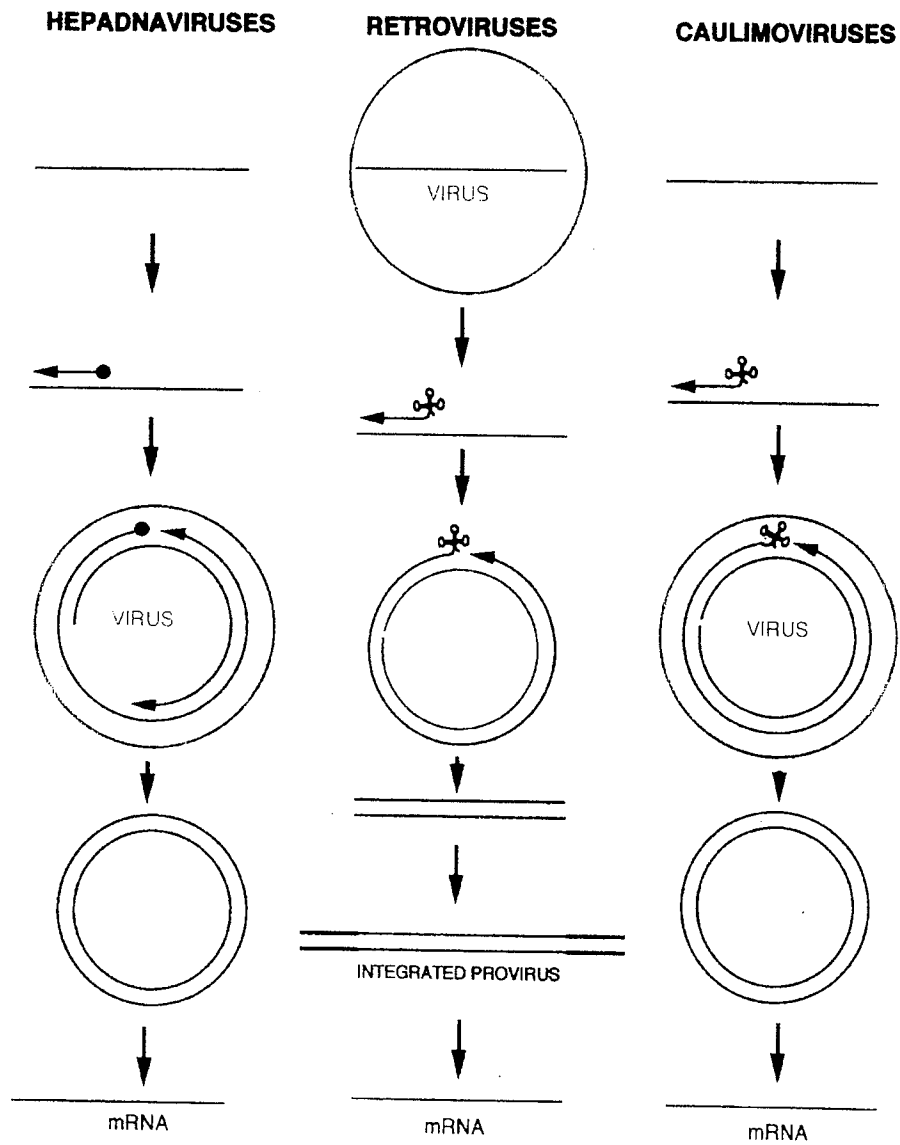Fig. 6. The life cycles of hepadnaviruses, retroviruses and caulimoviruses compared. DNA is shown in black, RNA in red and protein in green. Arrows refer to the direction of synthesis of DNA by reverse transcriptase.
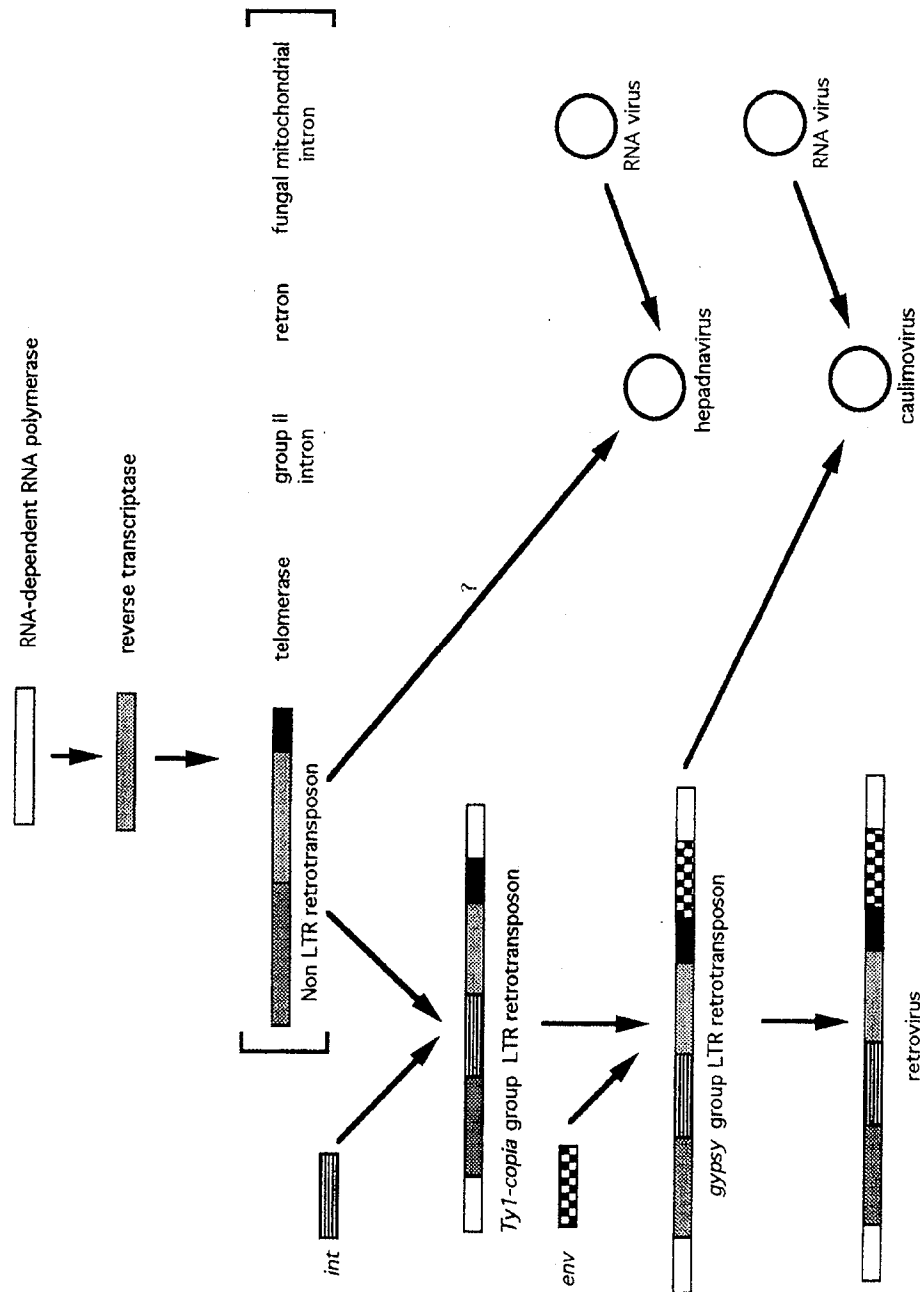
Fig. 7. A possible evolutionary history of retroelements. The genes are designated as in Fig. 5. The relative phylogeny of non-LTR retrotransposons, telomerases, group II introns and retrons is unsure.

mentioned here is still very unclear (Doolittle *et al.*, 1989; Xiong and Eickbush, 1990). Sequence comparisons between the reverse transcriptases argue that retrons and the fungal mitochondrial plasmids are grouped together, with group II introns being the nearest relatives to these sequences and non-LTR retrotransposons the next nearest. Additionally, some non-LTR retrotransposons transpose to the telomeric regions of *Drosophila* chromosomes (Biessman *et al.*, 1992; Levis *et al.*, 1994) suggesting an evolutionary link with telomerases. All this implies that the best candidate for the progenitor of all these elements belongs to an *ad hoc* group containing non-LTR retrotransposons, retrons, telomerases, fungal mitochondrial plasmids and group II introns, though it is impossible to say which, if any, came first.

Two properties unite virtually all retroelements, suggesting that they are fundamental to these elements and were present at their genesis. Firstly, virtually all reverse transcriptases prime the synthesis of new DNA from an RNA molecule with abundant secondary structure which is strongly associated with the enzyme. In the large majority of cases this is a tRNA. tRNAs themselves play a central role in living systems as the key component in the conversion of RNA to protein. RNA was probably the original genetic material and transfer RNAs and reverse transcriptase are the fundamental components of the machinery needed to synthesize DNA and protein, respectively, from RNA. The close association between the two in most retroelements to this day is striking and seems to this author a potent argument for this model of early evolution.

The second property uniting retroelements is their lack of any obvious advantage to their cellular hosts. With the single exception of telomerases, all reverse transcriptases are involved in the propagation of genetic elements which are not involved in the day-to-day functioning of cells. In fact, some retroelements are dangerous parasites (retroviruses, caulimoviruses and hepnaviruses). Thus, even though their origin probably lies at the dawn of cellular life, these elements have remained aloof from the business of enabling a cell to survive and replicate in an environment.

References

Augustin S., Müller M. W. and Schweyen R. J. (1990) Reverse self-splicing of group II intron RNAs *in vitro*. *Nature* **343**, 383–386.

Biessman H., Champion L. E., O'Hair M., Ikenaga K., Kasravi, B. and Mason J. M. (1992) Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. *EMBO J.* **11**, 4459–4469.

Bingham P. M. and Zachar Z. (1989) Retrotransposons and the FB element from *Drosophila melanogaster*. In *Mobile DNA* (Edited by Berg D. E. and Howe M. M.), pp. 485–502. American Society for Microbiology, Washington.

Blackburn E. H. (1992) Telomerases. *A. Rev. Biochem.* **61**, 113–129.

Boeke J (1989) Transposable elements in *Saccharomyces cerevisiae*. In *Mobile DNA* (Edited by Berg D. E. and Howe M. M.), pp. 335–374. American Society for Microbiology, Washington.

Bonneville J. M., Hohn T. and Pfeifer P. (1988) Reverse transcription in the plant virus cauliflower mosaic virus. In *RNA Genetics* (Edited by Domingo E., Holland J. J. and Ahlquist P.), Vol. I, pp. 23–42. CRC Press, Boca Raton, FL.

Carignani G., Groudinsky O., Frezza D., Schiavon E., Bergantino E. and Slonimski P. P. (1983) An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell* **35**, 733–742.

Cech T. R. and Bass B. L. (1986) Biological catalysis by RNA. *A. Rev. Biochem.* **55**, 599–629.

Darnell J. E. and Doolittle W. F. (1986) Speculations on the early course of evolution. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1271–1275.

Deragon J. M., Sinnett D. and Labuda D. (1990) Reverse transcriptase activity from human embryonal carcinoma cells NTera2D1. *EMBO J.* **9**, 3363–3368.

Doolittle W. F. and Sapienza C. (1980) Selfish genes, the phenotype, paradigm and genome evolution. *Nature* **284**, 601–603.

Doolittle R., Feng D. F., Johnson M. S. and McClure M. A. (1989) Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**, 1–29.

Finnegan D. J. (1989) The I factor and I-R hybrid dysgenesis in *Drosophila melanogaster*. In *Mobile DNA* (Edited by Berg D. and Howe M.), pp. 519–522. American Society for Microbiology, Washington DC.

Flavell A. J. and Smith D. B. (1992) A *Ty1-copia* group retrotransposon sequence in a vertebrate. *Molec. gen. Genet.* **233**, 322–326.

Flavell A. J., Jackson V., Iqbal M. and Riach I. (1994) *Ty1-copia* group retrotransposons in amphibian and reptile species. *Molec. Gen. Genet.* (in press).

Garfinkel D. J., Boeke J. D. and Fink G. R. (1985) Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42**, 507–517.

Gerlich W. H. and Robinson W. S. (1980) Hepatitis B virus contains protein attached to the 5′ terminus of its complete strand. *Cell* **21**, 801–809.

Hodgson C. P., Fisk R. Z., Arora P. and Chotani M. (1990) Nucleotide sequence of mouse virus-like

(VL30) retrotransposon BVL-1. *Nucleic Acids Res.* **18**, 673.

Hutchinson III C. A., Hardies S. C., Loeb D. D., Shehee W. R. and Edgell M. H. (1989) LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. In *Mobile DNA* (Edited by Berg D. and Howe M.), pp. 593–617. American Society for Microbiology, Washington DC.

Inouye S. and Inouye M. (1993) The retron: a bacterial retroelement required for the synthesis of msDNA. *Curr. Opinion genet. Dev.* **3**, 713–718.

Kim A., Terzian C., Santamaria P., Pelisson A., Prudhomme N. and Bucheton A. (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster. Proc. natn. Acad. Sci., U.S.A.* **91**, 1285–1289.

Lambowitz A. M. and Perlman P. S. (1990) Involvement of aminoacyl-tRNA synthetases and other proteins in group I and II intron splicing. *Trends biol. Sci.* **15**, 440–444.

Levis R. W., Ganesan R., Houtchens K., Tolar L. A. and Sheen F. (1994) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75**, 1083–1093.

Martin S. L. (1991) LINEs. *Curr. Opinion genet. Dev.* **1**, 505–508.

Martin S. L. and Branciforte D. (1993) Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Molec. cell. Biol.* **13**, 5383–5392.

Mizrokhi L. J., Georgieva S. V. and Ilyin Y. V. (1988) *jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from an internal promoter by RNA polymerase II. *Cell* **54**, 685–691.

Mueller M. W., Allmaier M., Eskes R. and Schweyen R. J. (1993) Transposition of Group II intron *aI1*

in yeast and invasion of mitochondrial genes at new locations. *Nature* **366**, 174–176.

Orgel L. E. and Crick F. H. C. (1980) Selfish DNA, the ultimate parasite. *Nature* **284**, 604–607.

Sapienza C. and Doolittle W. F. (1981) Genes are things you have whether you want them or not. *Cold Spring Harb. Symp. quant. Biol.* **45**, 177–182.

Schippan D. E. (1993) Telomeres and telomerases. *Curr. Opinion genet. Dev.* **3**, 759–763.

Sellem C. H., Lecellier G. and Belcour L. (1993) Transposition of a group II intron. *Nature* **366**, 176–178.

Shiba T. and Saigo K. (1983) Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster. Nature* **302**, 119–124.

Summers J. and Mason W. S. (1982) Replication of the genome on a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* **29**, 403–415.

Temin H. M. (1980) Origin of retroviruses from cellular movable genetic elements. *Cell* **21**, 599–600.

Temin H. M. (1989) Retrons in bacteria. *Nature* **339**, 254–255.

Tiollais P., Charnay A. and Vyas G. (1981) Biology of hepatitis B virus. *Science* **213**, 406–411.

Varmus H. E. and Brown P. (1989) Retroviruses. In *Mobile DNA* (Edited by Berg D. E. and Howe M. M.), pp. 53–108. American Society for Microbiology, Washington.

Wang H. and Lambowitz A. M. (1993) The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis *de novo* and may be related to reverse transcriptase and DNA polymerase progenitor. *Cell* **75**, 1071–1091.

Xiong Y. and Eickbush T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**, 3353–3362.

GENOME RESEARCH

The Unusual Phylogenetic Distribution of Retrotransposons: A Hypothesis

Jef D. Boeke

References	This article cites 59 articles, 25 of which can be accessed free at: http://www.genome.org/cgi/content/full/13/9/1975#References
	Article cited in: http://www.genome.org/cgi/content/full/13/9/1975#otherarticles
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes

To subscribe to *Genome Research* go to:
http://www.genome.org/subscriptions/

Review

The Unusual Phylogenetic Distribution of Retrotransposons: A Hypothesis

Jef D. Boeke

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Retrotransposons have proliferated extensively in eukaryotic lineages; the genomes of many animals and plants comprise 50% or more retrotransposon sequences by weight. There are several persuasive arguments that the enzymatic lynchpin of retrotransposon replication, reverse transcriptase (RT), is an ancient enzyme. Moreover, the direct progenitors of retrotransposons are thought to be mobile self-splicing introns that actively propagate themselves via reverse transcription, the group II introns, also known as retrointrons. Retrointrons are represented in modern genomes in very modest numbers, and thus far, only in certain eubacterial and organellar genomes. Archaeal genomes are nearly devoid of RT in any form. In this study, I propose a model to explain this unusual distribution, and rationalize it with the proposed ancient origin of the RT gene. A cap and tail hypothesis is proposed. By this hypothesis, the specialized terminal structures of eukaryotic mRNA provide the ideal molecular environment for the lengthening, evolution, and subsequent massive expansion of highly mobile retrotransposons, leading directly to the retrotransposon-cluttered structure that typifies modern metazoan genomes and the eventual emergence of retroviruses.

The Ancient Origin of Reverse Transcriptase

There are two arguments for an ancient origin of RT. The first is theoretical and is based on the now widely accepted proposal that an RNA world preceded the form of biology with which we are familiar, the DNA world. Darnell first articulated that RT must have been present during the time of the transition between these two worlds, and therefore, must be considered ancient (Darnell and Doolittle 1986; Fig. 1). The second argument is based on the fact that RT genes are very broadly distributed among the branches of the tree of life, and have largely (but not entirely) descended vertically by descent from an ancestral RT gene (Doolittle et al. 1989; Xiong and Eickbush 1990; Eickbush 1994; Malik et al. 1999). Furthermore, the RT gene has seemingly reinvented itself in multiple and diverse forms (Boeke and Stoye 1997). In addition to the familiar retroviruses, there are pararetroviruses, which package DNA, but replicate by reverse transcription, two major classes of retrotransposons (described in the following section), as well as a more bizarre group of elements found in bacteria and organellar genomes, and hence, referred to as the prokaryotic group. The discovery of RTs in bacteria, first in the form of msDNA (short for multicopy single-strand DNA) or retron elements (Yamanaka et al. 2002) and later in the form of retrointrons (Belfort et al. 2002), provided dramatic evidence in favor of an ancient origin for RT.

The highly diverse tree of retroelements can be rooted in the prokaryotic group of elements (Eickbush 1994). The prokaryotic group includes three types, that is, retrons, retroplasmids, and retrointrons. Retrons are RT genes that produce an unusual branched structure called msDNA made by reverse transcription of a precursor RNA primed from an internal guanosine residue—unlike other retroelements, they have no known function or ability to mobilize autonomously (Yamanaka et al. 2002). Thus far, they have been found only in a very limited subset of bacteria. Retroplasmids are known only from the mitochondria of certain fungi, replicate by reverse transcription, and exist in both circular and linear (hairpin) forms (Kuiper and Lambowitz 1988; Walther and Kennell 1999). The retrointrons, or group II introns, mobi-

lize or retrohome to empty target sites (unspliced versions of their host genes) via a very unconventional mechanism. The excised intron lariats insert into double-stranded target DNA (copies of the DNA containing the flanking exons but lacking the intron) by reversal of the normal splicing reaction, probably aided by the maturase activity of the RT proteins encoded by these elements. They are then converted into DNA by use of a target-primed reverse transcription (TPRT) mechanism similar to that used by non-LTR retrotransposons (Zimmerly et al. 1995a,b; Yang et al. 1996). Priming is facilitated by the action of a small endonuclease domain of the RT that cleaves the intact strand of the double-stranded target DNA.

Several independent arguments strongly suggest that the prokaryotic group of RT sequences is ancestral to the RT sequences of retrotransposons and retroviruses. Counterarguments to each of these proposals exist, but as a group, these proposals are compelling. (1) It is a simple evolutionary paradigm that things evolve progressively from a simple state to an ever more complex one. Retrons, retroplasmids, and retrointrons all encode a single RT protein, often with only that enzymatic activity, whereas retrotransposons and retroviruses always encode multiple enzyme activities and usually encode multiple separate proteins. These additional activities, which include proteases, zinc finger domains, at least three distinct types of endonucleases, and integrase, appear to have been recruited from eukaryotic host genomes at multiple times in evolution, probably using the same types of mechanisms used by retroviruses when they pick up cellular oncogenes (Telesnitsky and Goff 1997). A widely accepted extension of this simple argument is that the retroviruses and pararetroviruses evolved from LTR retrotransposons by acquiring new proteins conferring the ability to efficiently leave and re-enter host cells, also known as horizontal transfer or lateral transfer (Doolittle et al. 1989). (2) The RT of one member of the prokaryotic group has the ability to perform primer-independent synthesis, similar to RNA polymerase, the presumed ancestor of RT (Wang and Lambowitz 1993). (3) The RT sequences of the prokaryotic group are the most similar to the sequences of the presumed ancestral outgroup of sequences, the RNA-directed RNA polymerases (RdRPs). Non-LTR retrotransposons, LTR retrotransposons, and retroviral RTs are progressively less closely related to RdRP sequences (Eickbush 1994). (4)
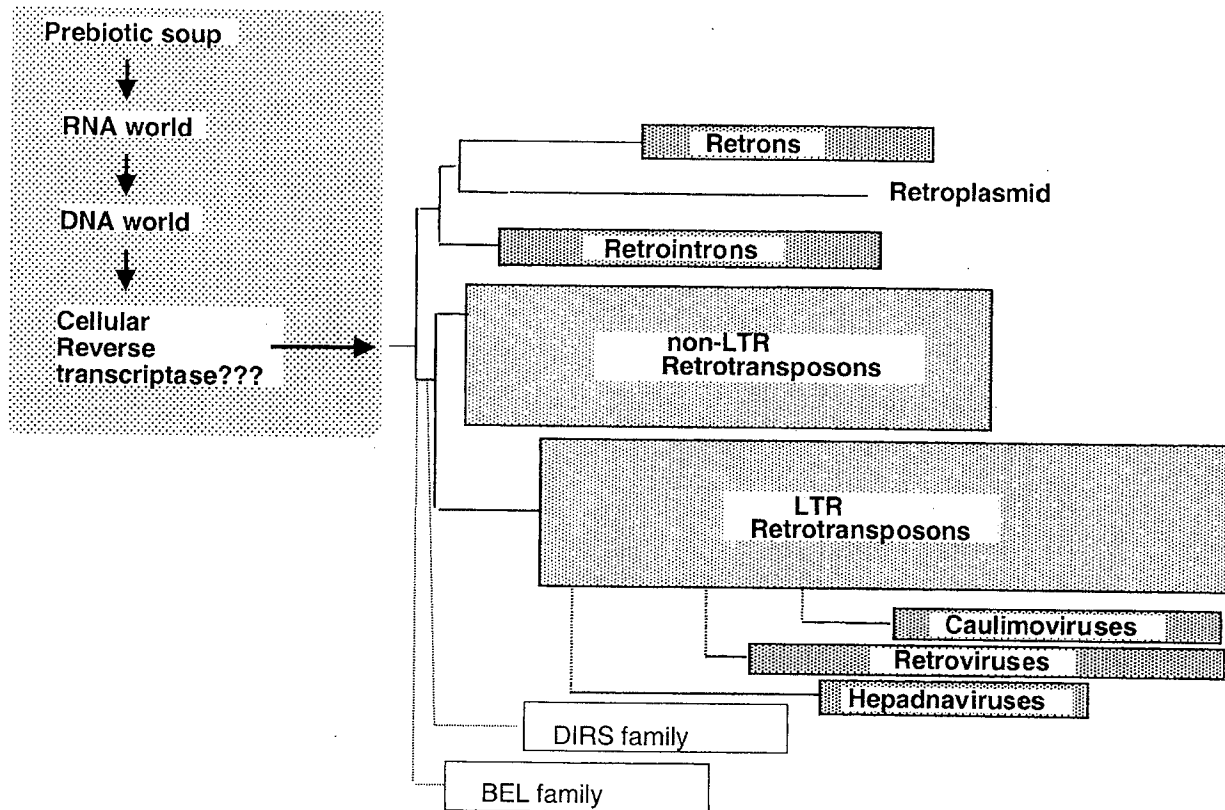
Figure 1 An ancient origin for reverse transcriptase. An early origin of a cellular reverse transcriptase is posited by the RNA world hypothesis (Darnell and Doolittle 1986) (*left*). The widespread existence of RT genes in prokaryotes, in eukaryotes, and their great molecular diversity (represented by the size of the boxes at the tip of each branch) also suggests an early origin for the RT gene. Tree diagram adapted from Eickbush (1994).

The sequence of telomerase, a specialized RT considered by many to represent an ancient eukaryotic enzyme, clusters with prokaryotic and non-LTR retrotransposon RT sequences (Eickbush 1997; Nakamura and Cech 1998).

Two Types of Retrotransposons That Mobilize by Distinct Mechanisms

The retrotransposons can be divided into two major groups, the non-LTR and the LTR retrotranposons. The mechanisms of these two types of retoelements are summarized briefly here and in Figure 2. In addition, two smaller retrotransposon families, the DIRS1 (Goodwin and Poulter 2001) and BEL (Malik et al. 2000) groups, appear to be distinct, but are much less widely distributed, and thus, will not be discussed further here.

Of the two major retrotransposon classes, the non-LTR retrotransposons, are less well understood mechanistically, but nevertheless, a good outline of the process exists (Kazazian and Moran 1998). The element mRNAs are translated in the cytoplasm, producing one or two proteins. One of these is a polyprotein with at least two critical activities, an endonuclease, and an RT. Most of the non-LTR elements also encode an RNA chaperone, whose role remains unclear. However, the endonuclease/RT protein is thought to bind the element RNA to form an RNP complex, which then enters the nucleus. This complex then acquires a host DNA target, in which a nick is made by the endonuclease. In a remarkable target-primed reverse transcription (TPRT) process, the 3' OH of the cleaved target DNA primes reverse transcription of the element RNA, at or near the 3' poly(A)

end. The mechanism of the cutting of the second strand and second-strand synthesis is less well understood, but may well be symmetrical with the first, involving a second round of TPRT with the newly made DNA strand serving as template.

LTR retrotransposons move via a mechanism quite similar to that used by retroviruses. Generally, two primary protein products are made, corresponding to retroviral Gag (coat proteins) and the readthrough product Gag-Pol (RT and other enzymes). The Gag proteins together with two RNA molecules are assembled into a virus-like particle (VLP). This encapsidation may serve to further protect the element's genomic RNA molecules from degradation. Reverse transcription occurs in the VLP, and is primed by a cellular tRNA (Chapman et al. 1992) or retrotransposon RNA fragment (Levin 1995). The initial product of the RT reaction, minus strand strong-stop DNA, is transferred to the 3' end of the RNA in a critical step that leads to subsequent completion of the minus strand DNA synthesis. If even relatively small amounts of RNA were lost exonucleolytically from the 5' or 3' end during this process, retrotransposition would fail. Several additional steps similar to those used by retroviruses, including a second priming event and strand transfer, lead to the final product of reverse transcription, a double-stranded DNA (Boeke and Stoye 1997; Telesnitsky and Goff 1997). RNA integrity is important for this process, which can take several hours to complete; however, a recombination-like template switching process can bypass damage to the element's RNA. The resulting DNA, together with the integrase protein (processed previously by an element-encoded protease from the RT precursor protein Gag-Pol)
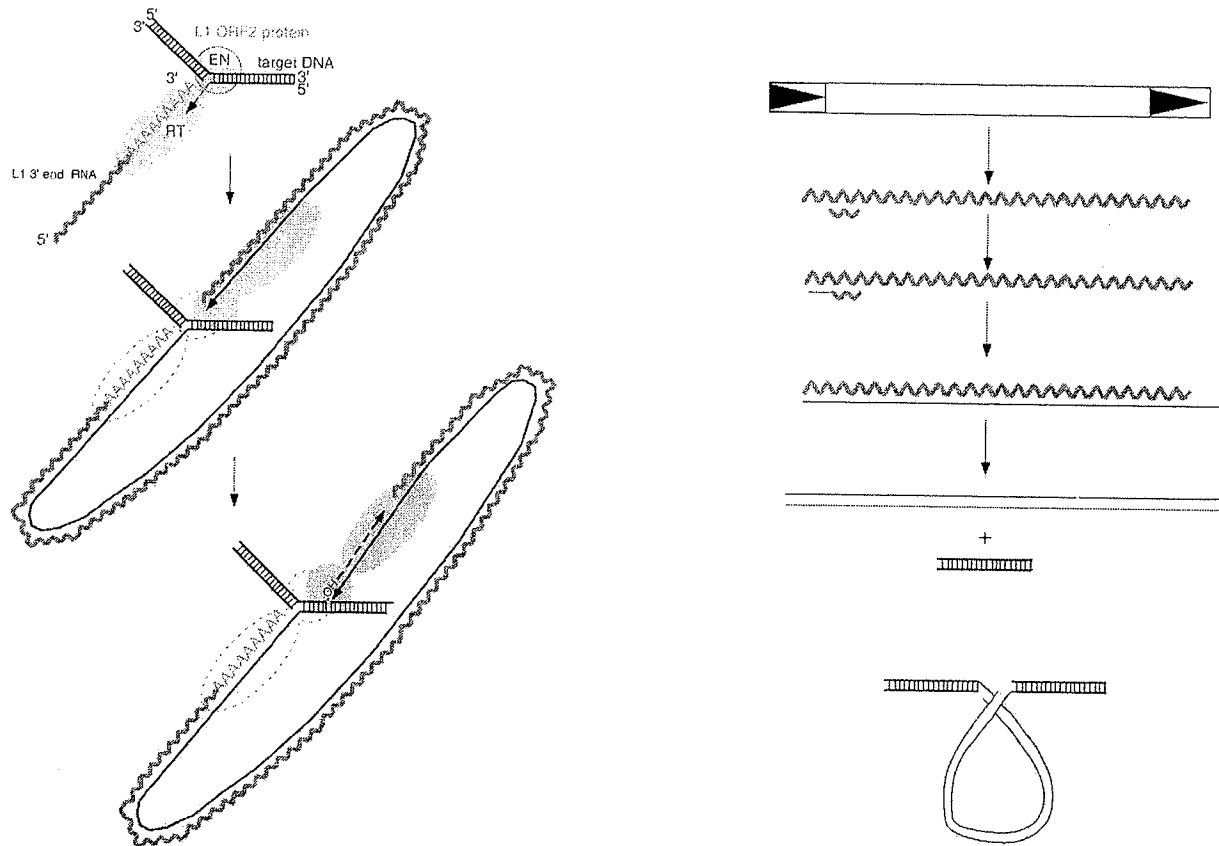
Figure 2 Retrotransposition mechanisms. The lifecycles of non-LTR (*left*) and LTR retrotransposons are outlined. Wavy lines are RNA molecules; thin black lines are cDNA strands.

is transported to the nucleus, where it inserts via a transesterification reaction very similar to that used by DNA transposons (Mizuuchi and Baker 2002).

Modern-Day Distribution of RT Genes

Faced with the assumption that RT is an ancient enzyme, it becomes difficult to explain the modern-day distribution of RT genes in the three kingdoms of life, Eubacteria, Archaea, and Eukarya. The majority (67%) of sequenced eubacterial species lack a detectable RT gene in their genome (Fig. 3). For those species of eubacteria that do contain RT genes, they mostly contain only one or two RT genes. The great majority of Archaea lack recognizable RTs altogether; the only exception to this trend, *Methanosarcina*, has a very large genome thought to have been formed by the incorporation of a large segment of a eubacterial genome as a late lateral transfer event in its evolution (Deppenmeier et al. 2002). This species contains a set of retrointrons similar to those found in eubacteria (Dai and Zimmerly 2003). In contrast, RT genes are found in virtually all eukaryotic genomes, and are generally found in 20 to >500,000 copies per genome. Even when adjusted for genome size, eukaryotes contain significantly more RT genes. Virtually all of these are non-LTR and/or LTR retrotransposons. In a recent grand synthesis, Bushman poetically described eukaryotic genomes as "genes floating on a sea of retrotransposons" (Bushman 2002), although an astute reviewer of this work points out that genes do not float, or else gene order colinearity would not be observed in genomic comparisons. Some well-known extreme examples of this include the

human genome, ~1,000,000 non-LTR retrotransposons, SINES, and endogenous retroviruses (Smit 1996) and the maize genome, estimated to contain ~200,000 copies of intact retrotransposons (SanMiguel et al. 1996; J. Bennetzen, pers. comm.). It is the abundance of retroelements that largely explains the C-value paradox in most metazoans. What led to such an abundance of RT genes?

It could be argued that the observed discrepancy is a simple consequence of genome streamlining in bacterial genomes. Although there is no doubt that streamlining is a major evolutionary force in both eubacteria and Archaea, one can consider as a control for the above conclusion the distribution of DNA transposons among the three kingdoms. DNA transposons are found in almost all eubacterial and archaeal genomes and typically are found between 10 and 100 copies. They are also found in eukaryotes, but have a somewhat spottier distribution there, being quite well represented in certain groups (*Drosophila, Caenorhabditis, Brassica*), but notably absent from others (*Saccharomyces, Schizosaccharomyces*).

The dramatic discrepancy in retroelement distribution between prokaryotes and eukaryotes strongly suggested to me that there was some special feature(s) of being eukaryotic that represented a permissive state for RT and allowed the evolution and proliferation of retrotransposons.

The Evolution of Eukaryotes and Their Retroelements

The release of numerous eubacterial, Archaeal, and eukaryotic genome sequences has provided extensive fodder for models of how eukaryotes evolved. It is clear that we eukaryotes contain a
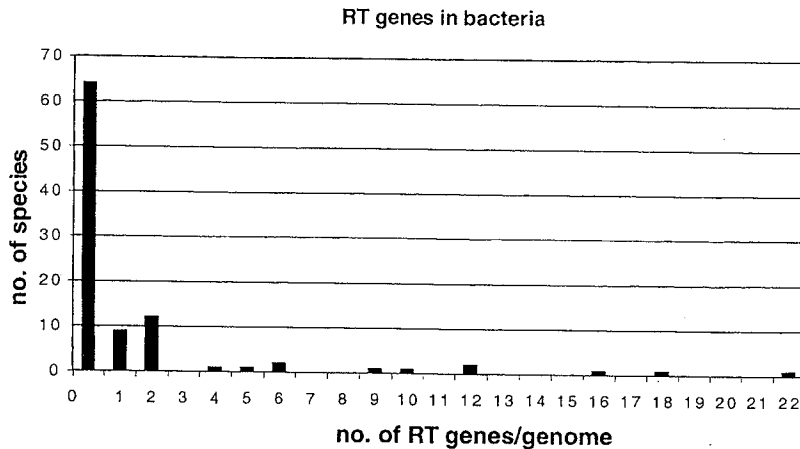
RT genes in bacteria

Figure 3 Bacterial genomes contain very few RT genes. A total of 96 completely sequenced bacterial genomes were searched by BLASTp on the comprehensive microbial resource at www. tigr.org. The two queries used were the LtrA RT from a *Lactobacillus lactis* group II intron (Q57005) and a retron RT from *Escherichia coli* (P23070). The number of BLAST hits with an E value <0.001 was tabulated for both queries, and the higher number was taken as the measure of RT gene number (visual inspection showed that this modestly inflated the number of RT genes as some of the low-scoring hits were false positives).

mixture of genes descended from Archaeal and eubacterial ancestor cells (Woese et al. 1990; Margulis 1996). The precise sequence of events in the evolution of eukaryotes has been debated hotly, but a consensus is developing about the major events that must have occurred. This consensus view will be recounted here briefly.

Archaea and eubacteria were two ancient lineages of cells that had evolved distinct mechanisms of transcription and DNA replication, among other things, but otherwise shared the fundamental properties of being unicellular heterotrophs. Symbiosis of eubacterial cells (the progenitor of the mitochondrion) and Archaeal cells ultimately led to a proto-eukaryote containing a eubacterial endosymbiont. This may have begun as a casual or accidental symbiosis, but at some point, provided some important selective advantage. Several other events followed, probably involving an additional cycle(s) of acquiring additional genomes via consumption (Taylor 1974), as well as the acquisition of a number of other distinctive eukaryotic features, which will be considered separately in the next section. These events gave rise to a primitive eukaryote with the recognizable nuclear genome and mitochondrial genome, each in a membrane-bounded compartment. Acquisition of an additional photosynthetic bacterium by consumption led to a plant lineage, but for simplicity, this will not be considered further here. Because the modern eubacteria contain RT genes and the Archaea largely lack them, I will make the fairly arbitrary assumption that the same was true at the dawn of eukaryotes. The eubacterially derived endosymbiont(s) slowly transferred its genes to the nucleus of the primitive eukaryotic cell, becoming ever more dependent on its host. Remarkably, this process of gene transfer from mitochondria to nucleus is functional in modern-day yeast cells, in which the transfer of mitochondrial gene segments to the nucleus can be observed experimentally (Thorsness et al. 2002). Through this process, RT genes present as retrointrons would be transferred readily to the nucleus through this passive and stochastic process. Movement of retrointrons via homing to near-cognate sites might well have led to a proliferation of introns and the evolution of the splicing apparatus as an intron-removal mechanism. The stage was set for the evolution of retrotransposons. What specific features of eukaryotic cells made this possible?

The Nuclear Membrane

The existence of a nuclear membrane would appear to be an impediment and not a help to the evolution of retrotransposition. The translation process occurs outside of the nucleus, whereas transposition happens inside, and therefore, retrotransposons have evolved transport mechanisms to overcome these barriers. Therefore, the existence of the nuclear membrane is inhibitory to successful retrotransposition.

Linear Chromosomes

The transition to linear chromosomes from the presumed ancestral circular state may well have provided an early opportunity for an RT gene to make itself indispensable to its host by acquiring the ability to lengthen telomeres, leading to the enzyme telomerase, still the major mechanism for telomere formation in modern eukaryotes (Nakamura and Cech 1998), providing an elegant solution to the end-replication problem posed by the termini of linear DNA molecules. However, this would provide a niche for but a single copy of RT. Also, a compelling case can be made that telomerase was a relatively late acquisition, by devolution of a retrotransposon (Pardue et al. 1997), as telomeres could, in principle, solve the end-replication problem via the formation of T-loops (Griffith et al. 1999). Thus, linear chromosomes per se do not provide a compelling opportunity for the evolution of retrotransposons.

Introns

As argued above, RT may have played an important role in the widespread accumulation of introns in primitive cells, although the timing of this event has also been the subject of much debate (Gilbert and Glynias 1993; Logsdon Jr. and Palmer 1994; Stoltzfus 1994; Logsdon Jr. 1998; Simpson et al. 2002). However, the simple existence of these introns did not confer any special selective advantage on RT genes. Rather, the proliferation of introns may well be a consequence of a permissive RNA environment that allowed them to mobilize more readily in the genome.

Sex and Diploidy

Donal Hickey (Hickey 1993) and Tim Bestor (Bestor 1999) have provided eloquent arguments that the evolution of sex and diploidy provides an opportunity for mobile elements to invade host species and march inexorably to fixation in the host genome, providing they do not decrease the fitness of their host >50%. However, this argument applies to both retrotransposons and DNA transposons, and thus, is insufficient to explain the selective amplification of retrotransposons in eukaryotes.

RNA Processing Machinery

The physical separation of the processes of transcription and translation and changes in gene organization (perhaps the consequence of the nascent eukaryal nuclear genome being bombarded with fragments of its endosymbiont guest DNA), and other factors, led to important changes in the way RNA was metabolized in eukaryotic cells. The major changes were the compartmentalization of single-coding regions (by and large) into stereotypical mRNA structures punctuated by a 5' cap structure and 3' poly(A) tail (Fig. 4). Not only are the structures of these mRNAs prominent uniquely in eukaryotic cells, but they also coordinate to play a critical role in eukaryotic translational ini-
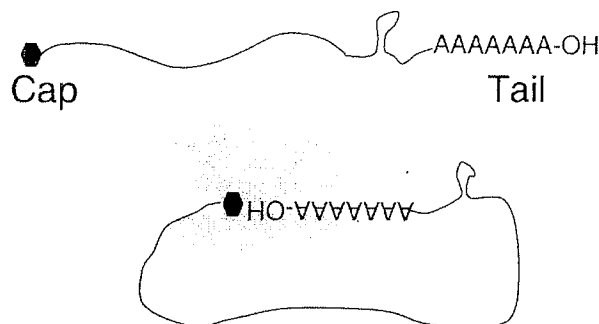
Figure 4 The cap and tail structure of eukaryotic mRNA.

tiation; the 5' cap and 3' poly(A) interact in the cytoplasm, effectively circularizing the RNA.

Other Factors

There are likely to be many factors that control the proliferation of transposable elements of all types in eukaryotes. For example, organisms such as yeasts and *Fugu*, with smaller genomes, tend to have much higher recombination rates, and these organisms carry lighter transposon burdens. Ergo, it can be argued that such high recombination rates are inconsistent with explosive types of transposon amplification, such as has been seen in humans and maize. Very high transposon copy numbers could cause extensive secondary damage to highly recombination-proficient genomes. Similarly, diverse mechanisms controlling the copy number of certain transposons' activity, such as cosuppression of Ty1 elements in yeast (Jiang 2002) and RNAi in many eukaryotes (Ketting et al. 1999), play important roles in controlling transposon copy numbers. Whereas such factors are undoubtedly of considerable importance in determining transposon copy numbers in individual species, they do not help to explain the general trend observed that eukaryotes tend to have very high retrotransposon copy numbers relative to prokaryotes.

The cap and tail hypothesis proposes that this unique terminal structure created three special molecular opportunities for the evolution of retrotransposons. First, these termini created a very stable long-lived genomic RNA freed from the necessity to be highly folded. Second, this RNA stability facilitated the recombinational acquisition of additional host gene modules needed for the formation of retrotransposons much more likely; the long mRNAs typical of retrotransposons and retroviruses were protected from destruction by exonucleases. Third, these terminal RNA structures provided precise punctuation marks defining the retrotransposon termini and facilitating their reproduction without the loss of even a single terminal nucleotide. These traits set the stage for the evolution of elaborate and precise processes of reverse transcription evolved by retrotransposons.

Why Did Eukaryotes Evolve Caps and Tails?

A number of theories have been advanced as to the evolution of the cap and tail. Extensive work on the molecular biology of translation has shown that the 5' cap and 3' tail structures are directly required for initiation of translation in eukaryotes. Additionally, both RNA structures are protective against terminal degradation of the RNA. In particular, the protective role of the 5' cap is revealed by the eukaryotic mRNA degradation pathway; this process occurs in three steps, (1) 3' deadenylation, leading to (2) decapping, followed by (3) 5' → 3' exonuclease action (Tucker and Parker 2000). Although 3' exonucleases are found in eukaryotic cells, they appear to play more specialized roles in mRNA

stability, such as nonstop decay and the destabilization of specific mRNAs (van Hoof and Parker 2002).

Polyadenylation occurs in all three kingdoms of life, although it only affects a subset of mRNAs in bacteria, and actually stimulates mRNA breakdown in prokaryotes (Steege 2000). Thus, certain components of the polyadenylation machinery predated the evolution of eukaryotes, and it appears that poly(A) simply acquired new functions in eukaryotes. In the literature and in discussions with colleagues, I've become aware of three theories regarding selective pressures leading to a need for a 5' cap. The first theory is that the compartmentalization of transcription leads to extensive opportunities for potentially inhibitory RNA folding prior to translation—potentially, such mRNA hairballs could occlude internal Shine Delgarno initiation sequences, whereas a terminal cap structure could more readily be recognized, like the end of a ball of yarn (Hershey and Merrick 2000). A second theory is that the complex nature of RNA processing in eukaryotes could lead to large numbers of misprocessed RNAs. Expression of inappropriately processed RNAs could lead to the expression of deleterious dominant negative protein fragments for example. Obviously, there are special pathways such as Nonsense-mediated decay (Frischmeyer and Dietz 1999; Gonzalez et al. 2001) and Nonstop decay, which deal with some of the RNA quality control issues raised by the existence of potentially inaccurate splicing machinery. However, a third type of proofreading is conferred by the obligatory circularization of mRNA during translational initiation—any RNA lacking intact 5' or 3' ends will not be translated (R. Green, pers. comm.).

Finally, Stewart Shuman has proposed that the cap arose to protect the RNA from 5' exonuclease action, and that the latter activity represented a type of primitive immunity against RNA viruses (Shuman 2002). Thus, the Xrn1p 5' exonuclease may have arisen in response to genomic RNA invaders, and the capping machinery evolved in parallel to protect endogenous cellular mRNAs. It is clear that eukaryotic cells evolved a series of different immunity mechanisms against invading RNA genomes, including the interferon system (Kumar and Carmichael 1998) and RNAi (Ketting et al. 1999). Needless to say, if this scenario is correct, the primitive immunity conferred by 5' exonuclease was quickly evaded by viruses that acquired caps by various nefarious means or evolved IRES elements that bypassed the cap requirement (Shuman 2002). Nevertheless, it would appear that the acquisition of the cap/5' exo strategy paradoxically set the stage for the evolution of a collection of internal genome invaders of eukaryotes, and eventually, retroviruses.

An interesting difference between bacteria and eukaryotes that may be related to differential RNA stability is the ability of eukaryotes to produce significantly longer proteins, such as the long polyproteins encoded by retrotransposons. Interestingly, a survey of bacterial genomes (Fig. 5) shows that bacteria, on average, encode shorter proteins than eukaryotes. This discrepancy becomes particularly acute when the longest ORFs are examined. The longest ORF in *Escherichia coli* K12, a putative invasin at 2383 codons, is less than half the length of the longest *Saccharamyoces cerevisiae* ORF, the *MDN1* gene at 4910 codons, and pales in comparison to human titin at 27,118 amino acids, encoded by an astonishingly long 82-kb mRNA (Labeit and Kolmerer 1995). This limit to ORF size does not represent an absolute expression block in bacteria, as some very large ORFs encoding nonribosomal polypeptide and polyketide synthases have been discovered in various bacterial species. It is possible that the simple lifestyle of prokaryotes generally requires shorter proteins than the complex lifestyle of eukaryotes. The evolution of a more stable mRNA structure in eukaryotes may well have contributed to the evolution of much greater potential protein structure complexity in general in eukaryotes.
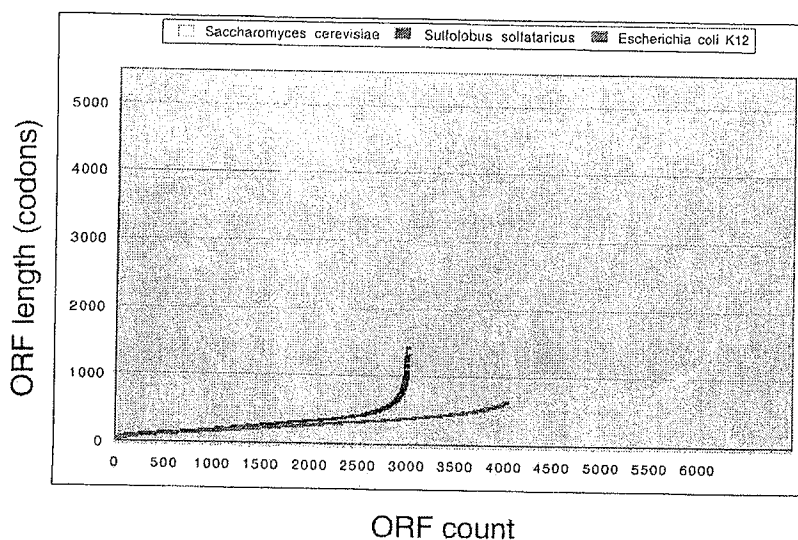
ORF count

Figure 5 Bacteria and Archaea encode smaller proteins than eukaryotes. The number of codons in each ORF for the indicated organisms was sorted in Excel, and a point was plotted for each protein. It can be seen readily that both the mean length and total length of the eukaryotic proteins are significantly higher than those of both the eubacterial and archaeal species. Results are typical (data not shown).

Bacteria Are RNA-Hostile

Recent work on the degradation of bacterial mRNAs has elucidated the basic molecular mechanisms, which are quite different from the eukaryotic mechanism (Table 1). In summary, eubacteria like *E. coli* degrade their RNAs through the combined effects of multiple endonucleases and 3′ exonucleases; many of the relevant activities are organized in degradosomes (Steege 2000). One of the well-studied endonucleases, RNAse E, nicks unstructured RNA regions adjacent to structured regions. Whereas the products of such nicking are not necessarily excellent direct substrates for the 3′ exonucleases, addition of a 3′ poly(A) tail creates an opportunity to initiate the degradation process by the degradosome; hence, mRNA polyadenylation leads to degradation in eubacteria.

If the cap and tail hypothesis is correct, it makes a number of predictions—for example, intact long RNA molecules should be difficult to detect in bacteria. It has long been known that it is extremely difficult to detect bacterial mRNAs by Northern blotting, and typical measurements of bacterial RNA half-lives range from seconds to minutes—far shorter than the half-lives of their eukaryotic counterparts, even when the mean mRNA half-life is adjusted for the cell generation time. (Fig. 6). Only a single value for average mRNA half-life is available from an Archaeal species, *Sulfolobus solfataricus*, which is among the slower-growing Archaea (some Archaea have fast doubling times similar to those of eubacteria), and its RNA half-life value is intermediate between eubacteria and *S. cerevisiae*, a eukaryote with a relatively short mRNA half-life (Bini et al. 2002). Examination of the mRNA degradation components encoded in eubacterial, Archaeal, and eukaryotic genomes shows that the eubacteria and Archaea share most of the same genes. Homologs of RNAse E, RNAse II, and polynucleotide phosphorylase are readily found by BLAST searching against Archaeal genomes. In contrast, Xrn1p homologs and capping enzyme homologs are absent from eubacteria and Archaea, but are common to all eukaryotes (Ananthara-man et al. 2002). Furthermore, like eubacteria, Archaea organize at least some of their genes in operons, and use Shine-Delgarno sequences to guide ribosomes to their initiation sites, at least in

some mRNAs (Shuman 2002), suggesting that translational initiation mechanisms in Archaea are more similar to those in eubacteria than in eukaryotes.

Eubacterial Retroelements Have Small, Highly Structured RNAs With Occluded 3′ Ends

A second prediction of the cap and tail model is that those retroelements that are found in eubacteria and Archaea will exhibit genomic features suggestive of protection against RNA degradation, such as short length, extensive secondary structure, and occluded 3′ ends. The two major classes of eubacterial retroelements display just these features. Retrointron RNA genomes are much shorter than retrotransposons and retroviruses, typically extending only 1–2 kb long versus 4–8 kb or more for typical retrotransposons and 10 kb or more for typical retroviruses. They are highly folded, their 5′ end is occluded via a 2′–5′ linkage and, moreover, they are always found in the form of a highly specific RNP, in which the RT-maturase protein is tightly bound to the intronic RNA. Importantly, the 3′ terminus of these molecules consists of a series of Watson Crick base pairs at the base of the domain VI stem of the intron, followed by two or three unpaired bases that can form a tertiary interaction with an internal segment in the intron (γ/γ′ sequences; Bonen and Vogel 2001). Similarly, the retron genome consists of a small, highly folded molecule in which the 3′ end of the RNA component is base paired to the 3′ end of the DNA component (Yamanaka et al. 2002).

Retrotransposon RNAs Are Capped and Polyadenylated

Nearly all retrotransposon RNAs contain caps and poly(A) tails, as do retroviral RNAs. The case is quite clear for LTR retrotransposons and retroviruses; there are many reports of poly(A) at the 3′ ends of LTR retrotransposon RNAs, and further evidence for posttranscriptionally added 3′ poly(A) tails in LTR retrotransposons can be found readily in EST databases. Capping is more laborious to evaluate, but some studies have been performed; for example, Ty1 mRNA was examined directly and found to be capped (Mules et al. 1998b), as are retroviral RNAs. Because LTR retrotransposons encode proteins required for their own mobility, and these must be translated from their mRNAs, it is extremely likely that all LTR-retrotransposon RNAs are capped.

One of the most important characteristics of non-LTR retrotransposons is that the vast majority of these elements actually encode poly(A) in their DNA. This 3′ poly(A) tract defines the element's 3′ end; many studies suggest that the 3′ poly(A) tract defines the site at which reverse transcription (TPRT) initiates (Moran et al. 1996). These poly(A) tails are peculiar in that they are apparently synthesized, at least in part, by RNA polymerase rather than the conventional polyadenylation machinery. However, it is possible that the 3′ poly(A) residues might be added post-transcriptionally using conventional polyadenylation. There are a few non-LTR retrotransposons such as the *Drosophila* I factor, which terminate not in poly(A), but in a related sequence, $(TAA)_n$. Clearly, the 3′ end of I factor RNA is not formed by conventional polyadenylation, but by transcription. Nevertheless, the number of TAA repeats can increase during retrotransposition, suggesting that a mechanism other than conven-

Table 1. RNA Degradation Components Found in the Three Kingdoms of Life

Enzyme	Activity	Distribution[a]
Prokaryotic		
RNAse E (*rne*)	Endoribonuclease; cuts single-stranded DNA adjacent to structured regions; provides internal access points for degradosome components	Eubacteria, Archaea (Eukarya—weak, exosome)
RNAse G (*rng*)	Endoribonuclease; cuts single-stranded DNA adjacent to structured regions; provides internal access points for degradosome components	Eubacteria, Archaea (weak) (Eukarya, weak)
RNAse III (*rnc*)	Endoribonuclease; cuts double-stranded DNA; provides internal access points for degradosome components	Eubacteria, Archaea some homologs in Eukarya
RNAse II (*rnb*)	3'–5' Exoribonuclease	Eubacteria
Polynucleotide Phosphorylase (*pnp*)	3'–5' Exoribonuclease	Eubacteria, Archaea some homologs in Eukarya
Oligoribonuclease (*orn*)	3'–5' Exoribonuclease	Eubacteria, Archaea
Eukaryotic		
Deadenylase (*CCR4*)	Removes nucleotides from 3' polyA during translation (Chen et al. 2002; Tucker et al. 2002)	Eukarya
Decapping enzyme (*DCP2*)	Removes 5' cap from deaenylated mRNA	Eukarya
Exonuclease (*XRN1*)	5'–3' exonuclease	Eukarya
Exosome (multiple)	3'–5' exonuclease	Eukarya (prokaryotes, weak)

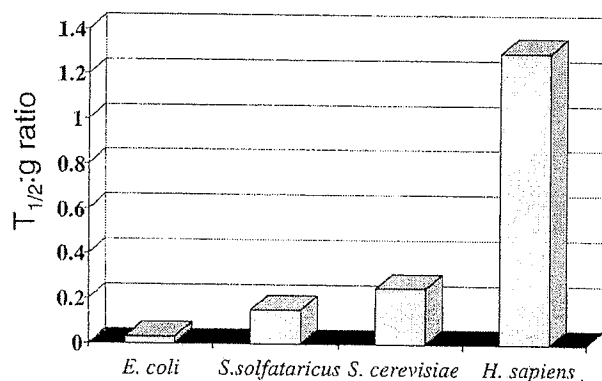[a]Taken from Anantharaman et al. (2002)

tional polyadenylation leads to the lengthening of the element 3' end, probably slippage by the I factor RT (Pritchard et al. 1988). Interestingly, the I factor 3' sequence can be replaced with poly(A), and the modified elements produce progeny elements with 3' poly(A) tails (Chambeyron et al. 2002). Intriguingly, a significant subset of human L1 elements carry a related (TAAA)$_n$ repeat in place of poly(A) (Szak et al. 2002). There are a few non-LTR retrotransposons, such as the CR1 element that terminate in a 3' terminal-repeated sequence unrelated to poly(A) (Burch et al. 1993). Presumably, these mRNAs have found another way to be circularized during translation, as they must be translated. Because this type of element lacking a poly(A)-like sequence is rare, I would propose that this is some late evolutionary adaptation. Clearly, the ancestral state of this family of elements is a 3' poly(A)tail.

Capping, however, has not been directly studied in the non-LTR retrotransposons, although the similarity of these elements' RNAs to mRNA strongly suggests that they are capped. There is evidence that the *Drosophila jockey* non-LTR retrotransposon is transcribed by RNA polymerase II, which is that its mRNA synthesis is α-amanitin sensitive (Mizrokhi et al. 1988). All known pol II mRNAs are capped, therefore, non-LTR mRNAs are unlikely

to be exceptions to this rule. Finally, the sequence of the human L1 element provides presumptive evidence for capping. Previous in vitro studies have shown that various RTs can readily copy the G residue comprising the cap, in spite of its unusual 5'–5' triphosphate linkage to the mRNA (Hirzmann et al. 1993; Volloch et al. 1995; Mules et al. 1998a). The L1 sequence starts with a run of a variable number G residues, which I propose has accumulated through multiple rounds of cap reverse transcription. In support of this exotic idea, the majority of extra single nucleotides accumulated at the 5' junction of experimentally isolated new full-length L1 insertions are G residues, whereas truncated L1 elements do not prefer single-G insertions (Symer et al. 2002; N. Gilbert, S.L. Lutz-Prigge, and J.V. Moran, pers. comm.).

A final exception to the general rule that eukaryotic retrotransposons are capped and polyadenylated is also instructive and supports the model, namely, the case of the *Alu* element and the related SINEs. These unusual elements don't need a cap, because they are not translated, but rely on retrotransposition proteins encoded by other non-LTR retrotransposons. Intriguingly, these elements are polyadenylated through transcription, even though they are transcribed by RNA polymerase III and, hence, are extremely unlikely to interact with the polyadenylation apparatus. However, these pol III transcripts lack a 5' cap. A different mechanism of protection from 5' exonuclease is adopted by these elements; as in the case of eubacterial retroelement 3' ends, *Alu* and related retroelements, as well as the tRNA-derived retroelements, are also highly folded and the 5' end of the RNA is always found in an extensively base-paired structure (e.g., see Sinnett et al 1992), which would protect it against XrnI-like 5' exonucleases.

Retrotransposon RNA levels are highly variable and tend to be tissue specific in metazoans, with high levels reached only in the germ line in most cases (Chaboissier et al. 1990; Branciforte and Martin 1994). Naturally, the abundance of retrotransposon RNAs is very strongly correlated with retrotransposition frequency. Because retrotransposition frequencies are set by some complex evolutionary interplay unique to each host/retrotransposon combination, it is not surprising that there is great variability in retrotransposon RNA levels. Nevertheless, there are some very dramatic cases of very high retrotransposon RNA lev-

Figure 6 Average mRNA half lives of diverse organisms, adjusted for generation time. The data are plotted directly from Bini et al. (2002).

els that provide strong evidence that the cap and tail structure are compatible with high levels of retrotransposon transcript stability. Of note, the *Drosophila* retrotransposon *copia* is so named because of its incredibly copious mRNA (Young and Hogness 1977), and yeast Ty1 mRNA levels are among the most abundant in the yeast cell (Curcio et al. 1990), with Ty1 mRNA visible as a discrete band in poly(A)-selected RNA preparations.

In conclusion, the stable and well-punctuated mRNA system was probably critical in allowing eukaryotes to evolve an ever more complex lifestyle, permitting longer more complex proteins and increased molecular diversity through alternative splicing. This same key change probably led to the extensive proliferation of retroelements, including retroviruses, in the many complex guises in which they are found today.

ACKNOWLEDGMENTS

REFERENCES

Anantharaman, V., Koonin, E.V., and Aravind, L. 2002. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30:** 1427–1464.

Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B., and Lambowitz, A.M. 2002. Mobile introns: Pathways and proteins. In *Mobile DNA II* (eds. N.L. Craig, et al.), pp. 761–783. American Society for Microbiology, Washington, D.C.

Bestor, T.H. 1999. Sex brings transposons and genomes into conflict. *Genetica* **107:** 289–295.

Bini, E., Dikshit, V., Dirksen, K., Drozda, M., and Blum, P. 2002. Stability of mRNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *RNA* **8:** 1129–1136.

Boeke, J.D. and Stoye, J.P. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In *Retroviruses* (eds. H. Varmus, et al.), pp. 343–435. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bonen, L. and Vogel, J. 2001. The ins and outs of group II introns. *Trends Genet.* **17:** 322–331.

Branciforte, D. and Martin, S.L. 1994. Developmental and cell-type specificity of LINE-1 expression in mouse testis: Implications for transposition. *Mol. Cell. Biol.* **14:** 2584–2592.

Burch, J.B., Davis, D.L., and Haas, N.B. 1993. Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc. Natl. Acad. Sci.* **90:** 8199–8203.

Bushman, F. 2002. *Lateral DNA transfer.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Chaboissier, M.C., Busseau, I., Prosser, J., Finnegan, D.J., and Bucheton, A. 1990. Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in *Drosophila melanogaster*. *EMBO J.* **9:** 3557–3563.

Chambeyron, S., Brun, C., Robin, S., Bucheton, A., and Busseau, I. 2002. Chimeric RNA transposition intermediates of the I factor produce precise retrotransposed copies. *Nucleic Acids Res.* **30:** 3387–3394.

Chapman, K.B., Byström, A.S., and Boeke, J.D. 1992. Initiator methionine tRNA is essential for Ty1 transposition in yeast. *Proc. Natl. Acad. Sci.* **89:** 3236–3240.

Curcio, M.J., Hedge, A.M., Boeke, J.D., and Garfinkel, D.J. 1990. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **220:** 213–221.

Dai, L. and Zimmerly, S. 2003. ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9:** 14–19.

Darnell, J.E. and Doolittle, W.F. 1986. Speculations on the early course of evolution. *Proc. Natl. Acad. Sci.* **83:** 1271–1275.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., et al. 2002. The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* **4:** 453–461.

Doolittle, R.F., Feng, D.F., Johnson, M.S., and McClure, M.A. 1989. Origins and evolutionary relationships of retroviruses. *Quart. Rev. Biol.* **64:** 1–30.

Eickbush, T.H. 1994. Origin and evolutionary relationships of retroelements. In *The evolutionary biology of viruses* (ed. S.S. Morse), pp. 121–157. Raven Press, Ltd., New York.
———. 1997. Telomerase and retrotransposons: Which came first? *Science* **277:** 911–912.

Frischmeyer, P.A. and Dietz, H.C. 1999. Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.* **8:** 1893–1900.

Gilbert, W. and Glynias, M. 1993. On the ancient nature of introns. *Gene* **135:** 137–144.

Gonzalez, C.I., Bhattacharya, A., Wang, W., and Peltz, S.W. 2001. Nonsense-mediated mRNA decay in *Saccharomyces cerevisiae*. *Gene* **274:** 15–25.

Goodwin, T.J. and Poulter, R.T. 2001. The DIRS1 group of retrotransposons. *Mol. Biol. Evol.* **18:** 2067–2082.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. 1999. Mammalian telomeres end in a large duplex loop. *Cell* **97:** 503–514.

Hershey, J.W.B. and Merrick, W.C. 2000. The pathway and mechanism of initiation of protein synthesis. In *Translational control of gene expression* (eds. N. Sonenburg, et al.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hickey, D.A. 1993. Molecular symbionts and the evolution of sex. *J. Hered.* **84:** 410–414.

Hirzmann, J., Luo, D., Hahnen, J., and Hobom, G. 1993. Determination of messenger RNA 5'-ends by reverse transcription of the cap structure. *Nucleic Acids Res.* **21:** 3597–3598.

Jiang, Y.W. 2002. Transcriptional cosuppression of yeast Ty1 retrotransposons. *Genes & Dev.* **16:** 467–478.

Kazazian Jr., H.H. and Moran, J.V. 1998. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19:** 19–24.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99:** 133–141.

Kuiper, M. and Lambowitz, A.M. 1988. A novel reverse transcriptase activity associated with mitochondrial plasmids of Neurospora. *Cell* **55:** 693–704.

Kumar, M. and Carmichael, G.G. 1998. Antisense RNA: Function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol. Mol. Biol. Rev.* **62:** 1415–1434.

Labeit, S. and Kolmerer, B. 1995. Titins: Giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270:** 293–296.

Levin, H.L. 1995. A novel mechanism of self-primed reverse transcription defines a new family of retroelements. *Mol. Cell. Biol.* **15:** 3310–3317.

Logsdon Jr., J.M. 1998. The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **8:** 637–648.

Logsdon Jr., J.M. and Palmer, J.D. 1994. Origin of introns—early or late? *Nature* **369:** 526; author reply 527–528.

Malik, H.S., Burke, W.D., and Eickbush, T.H. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16:** 793–805.

Malik, H.S., Henikoff, S., and Eickbush, T.H. 2000. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10:** 1307–1318.

Margulis, L. 1996. Archaeal-eubacterial mergers in the origin of Eukarya: Phylogenetic classification of life. *Proc. Natl. Acad. Sci.* **93:** 1071–1076.

Mizrokhi, L.J., Georgieva, S.G., and Ilyin, Y.V. 1988. *jockey*, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. *Cell* **54:** 685–691.

Mizuuchi, K. and Baker, T.A. 2002. Chemical mechanisms for mobilizing DNA. In *Mobile DNA II* (eds. N.L. Craig, R. Craigie, et al.), pp. 12–23. American Society for Microbiology, Washington, DC.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian Jr., H.H. 1996. High-frequency retrotransposition in cultured mammalian cells. *Cell* **87:** 917–927.

Mules, E.H., Uzun, O., and Gabriel, A. 1998a. In vivo Ty1 reverse transcription can generate replication intermediates with untidy ends. *J. Virol.* **72:** 6490–6503.
———. 1998b. Replication errors during in vivo Ty1 transposition are linked to heterogeneous RNase H cleavage sites. *Mol. Cell. Biol.* **18:** 1094–1104.

Nakamura, T.M. and Cech, T.R. 1998. Reversing time: Origin of telomerase. *Cell* **92:** 587–590.

Pardue, M.L., Danilevskaya, O.N., Traverse, K.L., and Lowenhaupt, K. 1997. Evolutionary links between telomeres and transposable elements. *Genetica* **100:** 73–84.

Pritchard, M.A., Dura, J.M., Pelisson, A., Bucheton, A., and Finnegan, D.J. 1988. A cloned I factor is fully functional in *Drosophila melanogaster*. *Mol. Gen. Genet.* **214:** 533–540.

SanMiguel, P., Tikhonov, A., Jin, Y., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P., Edwards, K., Lee, M., Avramova, Z.,

et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274:** 765–768.

Shuman, S. 2002. What messenger RNA capping tells us about eukaryotic evolution? *Nat. Rev. Mol. Cell. Biol.* **3:** 619–625.

Simpson, A.G., MacQuarrie, E.K., and Roger, A.J. 2002. Eukaryotic evolution: Early origin of canonical introns. *Nature* **419:** 270.

Sinnett, D., Richer, C., Deragon, J.M., and Labuda, D. 1992. Alu RNA transcripts in human embryonal carcinoma cells: Model of post-transcriptional selection of master sequences. *J. Mol. Biol.* **226:** 689–706.

Smit, A.F.A. 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6:** 743–748.

Steege, D.A. 2000. Emerging features of mRNA decay in bacteria. *RNA* **6:** 1079–1090.

Stoltzfus, A. 1994. Origin of introns—early or late. *Nature* **369:** 526–527; author reply 527–528.

Symer, D., Connelly, C., Szak, S., Caputo, E., Cost, G., Parmigiani, G., and Boeke, J. 2002. Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* **110:** 327–338.

Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D., and Boeke, J.D. 2002. Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **3:** research0052.

Taylor, F.J.R. 1974. Implications and extensions of the serial endosymbiosis theory of the origin of eukaryotes. *Taxon* **23:** 229–258.

Telesnitsky, A. and Goff, S.P. 1997. Reverse transcriptase and the generation of retroviral DNA. In *Retroviruses* (eds. H. Varmus, et al.), pp. 121–160. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Thorsness, M.K., White, K.H., and Thorsness, P.E. 2002. Migration of mtDNA into the nucleus. *Meth. Mol. Biol.* **197:** 177–186.

Tucker, M. and Parker, R. 2000. Mechanisms and control of mRNA decapping in *Saccharomyces cerevisiae. Annu. Rev. Biochem.* **69:** 571–595.

van Hoof, A. and Parker, R. 2002. Messenger RNA degradation: Beginning at the end. *Curr. Biol.* **12:** R285–287.

Volloch, V.Z., Schweitzer, B., and Rits, S. 1995. Transcription of the 5′-terminal cap nucleotide by RNA-dependent DNA polymerase:

Possible involvement in retroviral reverse transcription. *DNA Cell. Biol.* **14:** 991–996.

Walther, T.C. and Kennell, J.C. 1999. Linear mitochondrial plasmids of F. oxysporum are novel, telomere-like retroelements. *Mol. Cell* **4:** 229–238.

Wang, H. and Lambowitz, A.M. 1993. The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to the progenitor of reverse transcriptases and DNA polymerases. *Cell* **75:** 1071–1081.

Woese, C.R., Kandler, O., and Wheelis, M.L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87:** 4576–4579.

Xiong, Y. and Eickbush, T.H. 1990. Origin and evolution of retroelements based on their reverse transcriptase sequences. *EMBO J.* **9:** 3353–3362.

Yamanaka, K., Shimamoto, T., Inouye, S., and Inouye, M. 2002. Retrons. In *Mobile DNA II* (eds. N.L. Craig, et al.), pp. 784–795. American Society for Microbiology, Washington, DC.

Yang, J., Zimmerly, S., Perlman, P.S., and Lambowitz, A.M. 1996. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature* **381:** 332–335.

Young, M.W. and Hogness, D.S. 1977. A new approach for identifying and mapping structural genes in *Drosophila melanogaster*. In *Proceedings of the 1977 ICN/UCLA sumposium: Eucaryotic genetic systems* (eds. G. Wilcox, et al.), pp. 315–331. Academic Press, New York.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., and Lambowitz, A.M. 1995a. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83:** 529–538.

Zimmerly, S., Guo, H., Perlman, P.S., and Lambowitz, A.M. 1995b. Group II intron mobility occurs by target DNA -primed reverse transcription. *Cell* **82:** 545–554.

WEB SITE REFERENCES

www.tigr.org; Comprehensive Microbial Resource at the Institute for Genome Research.

mine. A single dose of clozapine increases dopamine release in the primate prefrontal cortex, and long-term administration increases basal extracellular dopamine concentration in the prefrontal cortex (21). Although this may not be the only mechanism by which clozapine elicits its effects on PCP-induced cognitive dysfunction, this activation of the dopamine system of the prefrontal cortex may contribute to the ability of clozapine to ameliorate the impairments in our model and, perhaps, in schizophrenia.

Our data show that repeated administration of PCP inhibits basal and stimulated dopaminergic function in the prefrontal cortex of the monkey brain. The deficiency of dopamine in the prefrontal cortex that is induced by repeated administration of PCP is associated with a long-lasting cognitive deficit, which is ameliorated by the atypical therapeutic drug clozapine. These effects are observed long after PCP administration is stopped and thus cannot be attributed to direct effects of the drug. This primate model of dopamine dysfunction in the cortex may provide a paradigm for investigating the pathophysiology underlying neuropsychiatric disorders associated with a primary cognitive dysfunction in the cortex and a dopaminergic deficit in the prefrontal cortex, as is hypothesized in schizophrenia (22). It also may provide a means for evaluating therapeutic agents that are selectively targeted toward alleviating cortical dopamine hypofunction.

REFERENCES AND NOTES

1. D. C. Javitt and S. R. Zukin, Am. J. Psychiatry 148, 1301 (1991).
2. A. S. Carlin, I. Grant, K. M. Adams, R. Reed, Am. J. Drug Alcohol Abuse 6, 273 (1979); J. E. Lewis and R. B. Hordan, in Phencyclidine: An Update, R. C. Petersen and R. C. Stillman, Eds. (National Institute of Drug Abuse, Rockville, MD, 1986), p. 66; J. Cosgrove and T. G. Newell, J. Clin. Psychol. 47, 159 (1991).
3. H. Klonhoff, H. C. Fibiger, G. H. Hutton, J. Nerv. Ment. Dis. 150, 291 (1970); S. Levin, J. Psychiatr. Res. 18, 27 (1984); T. E. Goldberg and J. M. Gold, in Psychopharmacology: The Fourth Generation of Progress, F. E. Bloom and D. J. Kupfer, Eds. (Raven, New York, 1995), p. 1245.
4. E. T. Fey, J. Consult. Psychol. 15, 311 (1951); T. Shallice, P. W. Burgess, C. D. Frith, Psychol. Med. 21, 661 (1991).
5. P. S. Goldman-Rakic, in Handbooks of Physiology, The Nervous System, vol. V, F. Plum and V. M. Mountcastle, Eds. (American Physiological Society, Bethesda, MD, 1987), p. 373.
6. B. Milner, Arch. Neurol. 9, 100 (1963); M. Freedman and M. Oscar-Berman, Behav. Neurosci. 100, 337 (1986); P. S. Goldman-Rakic, in Psychopathology and the Brain, B. J. Carroll and J. E. Barrett, Eds. (Raven, New York, 1991), p. 1; S. Park and P. S. Holzman, Arch. Gen. Psychiatry 49, 975 (1992); R. M. Ridley, B. A. Clark, L. J. Dunford, H. F. Baker, Neuroscience 52, 595 (1993); A. M. Owen, A. C. Roberts, J. R. Hodges, B. A. Summers, Brain 116, 1159 (1993).
7. M. Hertzman, R. C. Reba, E. V. Kotlyarove, Am. J. Psychiatry 147, 255 (1990); J. C. Wu, M. S. Buchs-

baum, W. E. Bunney, Jpn. J. Psychopharmacol. 11, 47 (1991).
8. B. G. Bunney, W. E. Bunney, A. Carlsson, in Psychopharmacology: The Fourth Generation of Progress, F. E. Bloom and D. J. Kupfer, Eds. (Raven, New York, 1995), p. 1205.
9. T. J. Brozoski, R. M. Brown, H. E. Rosvold, P. S. Goldman, Science 205, 929 (1979); T. Sawaguchi and P. S. Goldman-Rakic, ibid. 251, 947 (1991).
10. J. R. Taylor, J. D. Elsworth, R. H. Roth, J. R. Sladek, D. E. Redmond Jr., Brain 113, 617 (1990); J. R. Taylor, R. H. Roth, J. R. Sladek, D. E. Redmond Jr., Behav. Neurosci. 104, 564 (1990); J. D. Elsworth, A. Y. Deutch, D. E. Redmond Jr., J. R. Sladek, R. H. Roth, Neurosci. Lett. 114, 316 (1990).
11. Monkeys were trained and tested on the object retrieval with a detour task as described (10). Animals were treated with sterile saline (0.1 ml/kg twice a day, intramuscularly) or phencyclidine (0.3 mg/kg twice a day, intramuscularly) for 2 weeks. Two days after cessation of drug treatment, they began training on the task with an experimenter who was unaware of their drug treatment. Performance assessments and drug challenges were made no less than 7 days after cessation of drug treatment.
12. A. Diamond, in Thought Without Language, L. Weiskrantz, Ed. (Clarendon, Oxford, 1988); A. Diamond, S. Zola-Morgan, L. Squire, Behav. Neurosci. 102, 526 (1989).
13. J. D. Jentsch, J. D. Elsworth, D. E. Redmond Jr., R. H. Roth, J. Neurosci. 17, 1769 (1997).
14. J. D. Jentsch, A. Tran, D. Le, K. D. Youngren, R. H. Roth, Neuropsychopharmacology 17, 92 (1997).
15. B. Angrist, E. Peselow, M. Rubinstein, J. Corwin, J. Rotrosen, Psychopharmacology 78, 128 (1982); D. G. Daniel and K. F. Berman, J. Neuropsychiatry 1, 377 (1989); D. G. Daniel et al., J. Neurosci. 11, 1907 (1991); R. J. Dolan et al. Nature 378, 180 (1995).

16. P. Karlsson et al., Psychopharmacology 121, 309 (1995).
17. T. E. Goldberg et al., Br. J. Psychiatry 163, 43 (1993); T. E. Goldberg and D. R. Weinberger, J. Clin. Psychiatry 55 (suppl. B), 88 (1994).
18. J. Kane, G. Honigfeld, J. Singer, H. Y. Meltzer, Arch. Gen. Psychiatry 45, 789 (1988); M. A. Lee, P. A. Thompson, H. Y. Meltzer, J. Clin. Psychiatry 55 (no. 9) (suppl. B), 82 (1994).
19. Clozapine (1 mg/kg per day, intramuscularly) was administered for 3 consecutive days beginning on day 10 after cessation of PCP treatment. Performance was assessed before and 2 hours after final treatment with clozapine.
20. B. Saletu, J. Grunberger, L. Linzmayer, Pharmacopsychiatry 20, 12 (1987); B. L. Murphy, R. H. Roth, A. F. T. Arnsten, Neuropsychopharmacology 16, 433 (1997).
21. K. D. Youngren, B. Moghaddam, B. S. Bunney, R. H. Roth, Neurosci. Lett. 165, 41 (1994); K. D. Youngren et al., Soc. Neurosci. Abstr. 21, 1705 (1995).
22. D. R. Weinberger, K. F. Berman, B. P. Illowsky, Arch. Gen. Psychiatry 45, 609 (1988); K. L. Davis, R. S. Kahn, G. Ko, M. Davidson, Am. J. Psychiatry 148, 11 (1991); T. W. Robbins, Schizophr. Bull. 16, 391 (1990); A. Y. Deutch, J. Neural Transm. Suppl. 36, 61 (1992).
23. We thank P. S. Goldman-Rakic for critical reading of this manuscript; M. B. Bowers, B. S. Bunney, A. Y. Deutch, G. R. Heninger, and J. F. Tallman for helpful commentary; and A. Tran, Y. Pan, and the staff of the St. Kitts Biomedical Research Foundation for technical support. Supported in part by Public Health Service awards MH14092 and MH44866 (R.H.R.); by the Scottish Rite Schizophrenia Research Program NMJ, USA (J.D.J.); by RSA K05-MH00643 (D.E.R.); and by the Axion Research Foundation.

Telomerase Catalytic Subunit Homologs from Fission Yeast and Human

Toru M. Nakamura, Gregg B. Morin, Karen B. Chapman, Scott L. Weinrich, William H. Andrews, Joachim Lingner,* Calvin B. Harley, Thomas R. Cech†

Catalytic protein subunits of telomerase from the ciliate Euplotes aediculatus and the yeast Saccharomyces cerevisiae contain reverse transcriptase motifs. Here the homologous genes from the fission yeast Schizosaccharomyces pombe and human are identified. Disruption of the S. pombe gene resulted in telomere shortening and senescence, and expression of mRNA from the human gene correlated with telomerase activity in cell lines. Sequence comparisons placed the telomerase proteins in the reverse transcriptase family but revealed hallmarks that distinguish them from retroviral and retrotransposon relatives. Thus, the proposed telomerase catalytic subunits are phylogenetically conserved and represent a deep branch in the evolution of reverse transcriptases.

Telomerase is a ribonucleoprotein enzyme responsible in most eukaryotes for the complete replication of chromosome ends, or telomeres (1). Its RNA subunit provides the template for addition of short sequence repeats [typically 6 to 26 nucleotides (nts) to the chromosome 3' end (2). In ciliated protozoa and yeast, telomerase is regulated and the average telomere length is maintained (3). In most human somatic cells, however, telomerase activity cannot be detected and telomeres shorten with successive cell divisions (4). Telomerase activity reappears in immortalized cell lines and in about 85% of human tumors, which has led to studies of the usefulness of telomerase for cancer diagnostics and therapeutics (5, 6).

Telomerase RNA subunits have been identified and analyzed in ciliates, yeast, and mammals (2, 7), but the protein subunits have been elusive. In Tetrahymena, two telomerase-associated proteins (p80, p95) have been described (8), and p80 homologs have been found in humans and rodents (9); the presence of catalytic active site residues in these proteins has not been

established. Purification of telomerase from the ciliate *Euplotes aediculatus* yielded two proteins, p123 and p43 (*10*), that appear unrelated to p80 and p95 (*11*). p123 contains reverse transcriptase (RT) motifs and is homologous to yeast Est2 (Ever shorter telomeres) protein (*11*), which is essential for telomere maintenance in vivo (*12*). The RT motifs of Est2p are essential for telomeric DNA synthesis in vivo and in vitro (*11, 13*), supporting the conclusion that Est2p and p123 are the catalytic subunits of telomerase. The question remained whether there are two telomerases in biology, one based on p80- and p95-like proteins and one on p123/Est2p-like proteins.

To determine if Est2p/p123 is conserved among eukaryotes, we searched for homologs in the fission yeast *S. pombe* and humans. Polymerase chain reaction (PCR) amplification of *S. pombe* DNA was carried out with degenerate-sequence primers designed from the *Euplotes* p123 RT motifs B′ and C. Of the four prominent products generated, the ~120–base pair (bp) band encoded a peptide sequence homologous to p123 and Est2p. Using this PCR product as a probe for colony hybridization, we identified two overlapping clones from a genomic library and three from a cDNA library (*14*). None of the three cDNA clones was full length, so we used RT-PCR to obtain the NH_2-terminal sequences (*15*). This putative telomerase reverse transcriptase gene, *trt1*⁺, encoded a basic protein with a predicted molecular mass of 116 kilodaltons (kD) (Fig. 1A). The sequence similarity to p123 and Est2p was especially high in the seven RT motifs (Table 1) and in motif T (Telomerase-specific) (Fig. 2). Fifteen introns, ranging from 36 to 71 bp, interrupted the coding sequence. All had consensus splice and branch site sequences (*16*).

If *trt1*⁺ encodes the telomerase catalytic subunit in *S. pombe*, deletion of the gene would be expected to result in telomere shortening and perhaps cellular senescence as seen with the *est2* mutants in *S. cerevisiae* (*11, 13*). To test this, we created two deletion constructs (Fig. 1A), one removing motifs B′ through E in the RT domain, and the second deleting 99% of the open reading frame (ORF). Haploid cells grown from both types of spores showed progressive telomere shortening to the point where hybridization to telomeric repeats became al-

most undetectable (Fig. 1B). Senescence was indicated by (i) reduced ability of the cells to grow on agar, typically by the fourth streak-out after germination; (ii) the appearance of colonies with increasingly ragged edges (Fig. 1C); and (iii) the increasing fraction of elongated cells (Fig. 1D). When individual enlarged cells were separated on the dissecting microscope, the majority underwent no further division. The same *trt1*⁻ cell population always contained normal-size cells that continued to divide but frequently produced nondividing progeny. The telomerase-negative survivors may use a recombinational mode of telomere maintenance as documented in budding yeast strains with deletions of telomere-replication genes (*12, 17*).

A candidate human p123/Est2p/Trt1p homolog was identified by a BLAST search of the EST (expressed sequence tag) database (GenBank AA281296). This EST was the top-ranked match in sequence searches with *Euplotes* p123 ($P = 3.3 \times 10^{-6}$) and *S. pombe* Trt1p ($P = 9.7 \times 10^{-7}$). The human EST was not found in searches with yeast

Fig. 1. The gene for the *S. pombe* telomerase protein and phenotypes associated with its deletion. (**A**) The *trt1*⁺ locus, the location of the ~120 bp PCR product that led to its identification, and the regions replaced by *ura4*⁺ or *his3*⁺ genes in the *trt1*⁻ mutants (K, Kpn I; Xb, Xba I; H, Hind III; Xc, Xca I; Xh, Xho I). (**B**) Telomere shortening phenotype of *trt1*⁻ mutants. A *trt1*⁺/*trt1*⁻ diploid (*28*) was sporulated and the resulting tetrads were dissected and germinated on a YES (Yeast Extract medium Supplemented with amino acids) plate (*29*). Colonies derived from each spore were grown at 32°C for 3 days, and streaked successively to fresh YES plates every 3 days. A colony from each round was placed in 6 ml of YES liquid culture at 32°C and grown to stationary phase, and genomic DNA was prepared. After digestion with Apa I, DNA was subjected to electrophoresis on a 2.3% agarose gel, stained with ethidium bromide to confirm approximately equal loading in each lane, transferred to a nylon membrane, and hybridized to a telomeric DNA probe. The Apa I site is located 30 to 40 bp away from telomeric repeat sequences in chromosomes I and II. (**C**) Colony morphology of *trt1*⁺ and *trt1*⁻ cells. Cells plated on MM [Minimal Medium (*29*) with glutamic acid substituted for NH_4Cl] were grown for 2 days at 32°C prior to photography. (**D**) Micrographs of *trt1*⁺ and *trt1*⁻ cells grown as in (C).

T. M. Nakamura, J. Lingner, T. R. Cech, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309–0215, USA.

G. B. Morin, K. B. Chapman, S. L. Weinrich, W. H. Andrews, C. B. Harley, Geron Corporation, 230 Constitution Drive, Menlo Park, CA 94025, USA.

*Present address: Swiss Institute for Experimental Cancer Research, 1066 Epalinges/VD, Switzerland.
†To whom correspondence should be addressed.

Est2p, but subsequent pairwise comparison of these sequences showed a convincing match. Sequencing of the rest of the cDNA clone containing the EST revealed all eight TRT (Telomerase Reverse Transcriptase) motifs, but not in a single ORF (18). We used the sequence information from this incomplete cDNA clone to isolate an extended cDNA clone from a library of 293 cells, an adenovirus E1-transformed human embryonic kidney cell line (19). This cDNA clone (pGRN121) had a 182-bp insert relative to the EST clone, which increased the spacing between motifs A and B' (18) and put all seven RT motifs and the telomerase-specific motif T motifs in a single contiguous ORF (Fig. 2).

RT-PCR amplification of RNA from 293 cells and from testis each gave two products differing by 182 bp (20). The larger and smaller products from testis RNA were sequenced and found to correspond exactly to pGRN121 and the EST cDNA, respectively.

The relative abundance of hTRT mRNA was assessed in six telomerase-negative mortal cell strains and six telomerase-positive immortal cell lines (21) (Fig. 3). The steady-state level of hTRT mRNA was higher in immortal cell lines with active telomerase (6) than in any of the telomerase-negative cell strains tested. Telomerase activity was more strongly correlated with the abundance of hTRT mRNA than with that of telomerase RNA

(hTR) (7). In contrast, the abundance of mRNA for the human p80 homolog TP1 (9) did not correlate with telomerase activity (Fig. 3). Thus, while our proposal that hTRT is the catalytic subunit of human telomerase is based mainly on protein structural features

Table 1. Amino acid sequence identity between telomerase reverse transcriptases. Each value is % identity (% similarity in parentheses) based on RT motifs 1, 2, and A through E in Fig 2C.

	hTRT	SpTrt1p	Est2p
Ea p123	26 (49)	28 (45)	24 (46)
Est2p	25 (46)	27 (48)	
SpTrt1p	30 (47)		

Fig. 2. Structure and RT sequence motifs of telomerase proteins. **(A)** Locations of telomerase-specific motif T and conserved RT motifs 1, 2, and A through E (24) are indicated by colored boxes. The open rectangle labeled HIV-1 (Human Immunodeficiency Virus) RT delineates the portion of this protein shown in (B). pl, isoelectric point. **(B)** The crystal structure of the p66 subunit of HIV-1 RT (Brookhaven code 1HNV). Color-coding of RT motifs matches that in (A). The view is from the back of the right hand, which allows all motifs to be seen. **(C)** Multiple sequence alignment of telomerase RTs and members of other RT families (Sc_al, cytochrome oxidase group II intron 1-encoded protein from S. cerevisiae mitochondria; Dm_TART, reverse transcriptase from Drosophila melanogaster TART non-LTR retrotransposable element). Boldface residues indicate identity of at least three telomerase sequences in the alignment. Colored residues are highly con-

served in all RTs and shown as space-filled residues in (B). The number of amino acids between adjacent motifs or to the end of the polypeptide is indicated. TRT con and RT con, consensus sequences for telomerase RTs (this work) and non-telomerase RTs (24) (amino acids are designated h, hydrophobic, A, L, I, V, P, F, W, M; p, polar, G, S, T, Y, C, N, Q; c, charged, D, E, H, K, R). Red arrowheads show some of the systematic differences between telomerase proteins and other RTs. Red rectangle below motif E highlights the primer grip region discussed in the text. Abbreviations for the amino acids are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. The nucleotide sequences of the S. pombe trt1' gene and the human TRT cDNA (pGRN121) have been deposited in GenBank (accession nos. AF015783 and AF015950, respectively).

(similarity of RT motifs, the T motif, molecular mass > 100 kD, pI > 10), the correlation of its mRNA expression level with activity also supports this conclusion.

Sequence alignment of the four telomerase genes revealed features similar to other reverse transcriptases, as well as differences that serve as hallmarks of the telomerase subgroup. The new T motif is one telomerase-specific feature not found in the other RTs examined. Another is the distance between motifs A and B′, which is
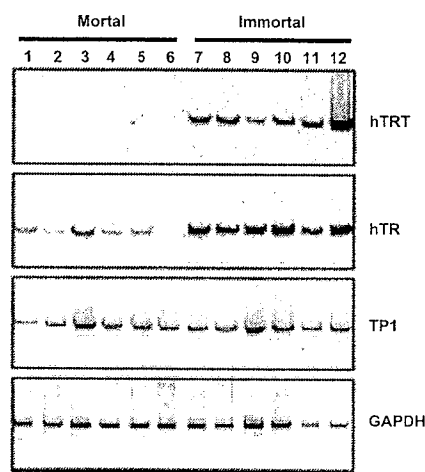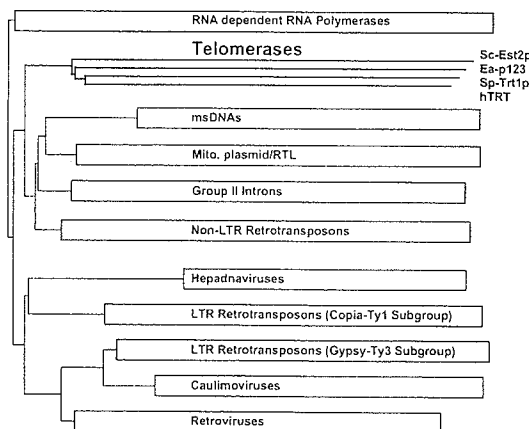

Fig. 3. Expression of hTRT in telomerase-negative mortal cell strains (lanes 1 to 6) and telomerase-positive immortal cell lines (lanes 7 to 12). RT-PCR (21) for hTRT, hTR (human telomerase RNA component), TP1 (telomerase-associated protein related to *Tetrahymena* p80), and GAPDH (to normalize for equal amounts of RNA template) was carried out on RNA from: (1) human fetal lung fibroblasts GFL, (2) human fetal skin fibroblasts GFS, (3) adult prostate stromal fibroblasts 31YO, (4) human fetal knee synovial fibroblasts HSF, (5) neonatal foreskin fibroblasts BJ, (6) human fetal lung fibroblasts IMR90, (7) melanoma LOX IMVI (8) leukemia U251, (9) NCI H23 lung carcinoma, (10) colon adenocarcinoma SW620, (11) breast tumor MCF7, and (12) human 293 cells.

Fig. 4. A possible phylogenetic tree of telomerases and retroelements rooted with RNA-dependent RNA polymerases. After sequence alignment of motifs 1, 2, and A through E (178 positions, Fig. 2C) from four TRTs, 67 RTs, and three RNA polymerases, the tree was constructed using the Neighbor Joining method (30). Elements from the same class that are located on the same branch of the tree are simplified as a box. The length of each box corresponds to the most divergent element within that box.

longer in the TRTs than in other RTs (Fig. 2A). These amino acids can be accommodated as an insertion within the "fingers" region of the structure that resembles a right hand (22, 23) (Fig. 2B). Within the motifs, there are a number of substitutions of amino acids (red arrowheads in Fig. 2C) that are highly conserved among the other RTs. For example, in motif C the two aspartic acid residues (DD) that coordinate active site metal ions (22) occur in the context hxDD(F/Y) in the telomerase RTs compared to (F/Y)xDDh in the other RTs (24). Another systematic change characteristic of the telomerase subgroup occurs in motif E, where WxGxSx appears to be the consensus among the telomerase proteins, whereas hLGxxh is characteristic of other RTs (24). This motif E is called the "primer grip" (23), and mutations in this region affect RNA priming but not DNA priming (25). Because telomerase uses a DNA primer, the chromosome 3′ end, it is not unexpected that it should differ from other RTs in this region. Given that the simple change from Mg^{2+} to Mn^{2+} allows HIV RT to copy a small region of a template in a repetitive manner (26), it is tempting to speculate that some of the distinguishing amino acids in the TRTs may cause telomerase to catalyze repetitive copying of the template sequence of its tightly bound RNA subunit.

Using the seven RT domains (Fig. 2C) defined by Xiong and Eickbush (24), we constructed a phylogenetic tree that includes the four telomerase RTs (Fig. 4). The TRTs appear to be more closely related to RTs associated with msDNA (multicopy single-stranded DNA), group II introns, and non-LTR (Long Terminal Repeat) retrotransposons than to the LTR-retrotransposon and viral RTs. The relationship of the telomerase RTs to the non-LTR branch of retroelements is intriguing, given that the latter elements have replaced telomerase for telomere maintenance in *Drosophila*

(27). However, the most striking finding is that the TRTs form a discrete subgroup, about as closely related to the RNA-dependent RNA polymerases of plus-stranded RNA viruses such as poliovirus as to retroviral RTs. In view of the fact that the four telomerase genes are from evolutionarily distant organisms—protozoan, fungi, and mammals—this separate grouping cannot be explained by lack of phylogenetic diversity in the data set. Instead, this deep branching suggests that the telomerase RTs are an ancient group, perhaps originating with the first eukaryote.

The primary sequence of hTRT and eventual reconstitution of active human telomerase may be used to discover telomerase inhibitors, which in turn will permit additional testing of the anti-tumor effects of telomerase inhibition. The correlation between hTRT mRNA levels and human telomerase activity shown here indicates that hTRT also has promise for cancer diagnosis. With an essential protein component of telomerase now in hand, the stage is set for more detailed investigation of fundamental and applied aspects of this ribonucleoprotein enzyme.

REFERENCES AND NOTES

1. C. W. Greider and E. H. Blackburn, *Nature* 337, 331 (1989).
2. C. W. Greider, K. Collins, C. Autexier, in *DNA Replication in Eukaryotic Cells*, M. L. DePamphilis, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996) pp. 619–638.
3. L. A. Klobutcher, M. T. Swanton, P. Donini, D. M. Prescott, *Proc. Natl. Acad. Sci. U.S.A.* 78, 3015 (1981); M. J. McEachern and E. H. Blackburn, *Nature* 376, 403 (1995); J. P. Cooper, E. R. Nimmo, R. C. Allshire, T. R. Cech, *ibid.* 385, 744 (1997); S. Marcand, E. Gilson, D. Shore, *Science* 275, 986 (1997); V. A. Zakian, *ibid.* 270, 1601 (1995).
4. C. B. Harley, A. B. Futcher, C. W. Greider, *Nature* 345, 458 (1990); N. D. Hastie *et al.*, *ibid.* 346, 866 (1990); C. B. Harley *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 59, 307 (1994).
5. C. M. Counter *et al.*, *EMBO J.* 11, 1921 (1992); J. W. Shay and S. Bacchetti, *Eur. J. Cancer* 33, 787 (1997).
6. N. W. Kim *et al.*, *Science* 266, 2011 (1994).
7. J. Feng *et al.*, *ibid.* 269, 1236 (1995).
8. K. Collins, R. Kobayashi, C. W. Greider, *Cell* 81, 677 (1995).
9. L. Harrington *et al.*, *Science* 275, 973 (1997); J. Nakayama, M. Saito, H. Nakamura, A. Matsuura, F. Ishikawa, *Cell* 88, 875 (1997).
10. J. Lingner and T. R. Cech, *Proc. Natl. Acad. Sci. U.S.A.* 93, 10712 (1996).
11. J. Lingner *et al.*, *Science* 276, 561 (1997).
12. T. S. Lendvay, D. K. Morris, J. Sah, B. Balasubramanian, V. Lundblad, *Genetics* 144, 1399 (1996).
13. C. M. Counter, M. Meyerson, E. N. Eaton, R. A. Weinberg, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
14. PCR primers were based on motif B′ (YARACHAARGGHATYCCHYARGG) and motif C (NGTNATDARDARRTARTCRTC), where D = G, A, or T; H = T, A, or C; Y = T or C; R = A or G; and N = G, A, T or C. Approximately 10^6 colonies of an *S. pombe* Hind III–cut genomic library in pWH5 [A. Wright, K. Maundrell, W. D. Heyer, D. Beach, P. Nurse, *Plasmid* 15, 156 (1986)] and 10^5 colonies of an *S. pombe* cDNA library [J.-P. Javerzat, G. Cranston, R. C. Allshire, *Nucleic Acids Res.* 24, 4676 (1996)] were screened.

15. These cDNA clones terminated in the exon flanked by introns 4 and 5. Poly A⁺ RNA from *S. pombe* was reverse transcribed using primer M2-B14 (CCTTG-GAAAAATCCATTGAAGCCACATGTG). The resulting cDNA was ligated to oligonucleotide pGGGCCGT-GTTGGCCTAGTTCTCTGCTCddA using T4 RNA ligase and amplified by two rounds of PCR: in the first round, we used primers M2-B14 and Adapt-Sfi (GAGGAGGAGAAGAGCAGAGAACTAGGCCAACA-CGGCCC), and in the second, we used primers M2-B15 (AAAGTGGTATGCCAGAAATCTGAAGG-TAAT) and Adapt-Sfi.
16. M. Q. Zhang and T. G. Marr, *Nucleic Acids Res.* 22, 1750 (1994). Sequencing of the three original cDNA clones and nine cloned RT-PCR products revealed that two 3′ splice sites were used at similar frequency for splicing of intron 8. These sites were separated by 3 nts, giving rise to predicted proteins of 988 and 989 amino acids.
17. V. Lundblad and E. H. Blackburn, *Cell* 73, 347 (1993); M. J. McEachern and E. H. Blackburn, *Genes Dev.* 10, 1822 (1996).
18. Clone 712562 was obtained from the I.M.A.G.E. (Integrated Molecular Analysis of Genomes and Their Expression) Consortium [G. Lennon, C. Auffray, M. Polymeropoulos, M. B. Soares, *Genomics* 33, 151 (1996)]. This clone did not encode a contiguous portion of a TRT because motifs B′, C, D, and E were contained in a different ORF than the more NH₂-terminal motifs. In addition, the distance between motifs A and B′ was substantially shorter than that of the other three TRTs.
19. A lambda cDNA library from the human 293 cell line, which has high levels of telomerase activity, was partitioned into 25 pools containing ~200,000 plaques each. These were screened by PCR with primers LT5 (CGGAAGAGTGTCTGGAGCAA) and LT6 (GGATGAAGCGGAGTCTGGA) to the 5′ region of the #712562 clone insert. Six subpools of one positive primary pool were further screened by PCR. One positive subpool was then screened by plaque hybridization with a probe from the 5′ region of clone #712562. One phage was positively identified and the ~4 kbp insert from this clone was excised and subcloned into the pBluescript II SK+ vector (Stratagene) as an Eco RI fragment.
20. Polyadenylated RNAs from human testis and from the 293 cell line were amplified using a nested PCR strategy. The first primer set was TCP1.1 (GTGAAG-GCACTGTTCAGCG) and TCP1.15 (CGCGTGGGT-GAGGTGAGGTG); the second primer set was TCP1.14 (CTGTGCTGGGCCTGGACGATA) and bTCP6 (AGCTTGTTCTCCATGTCGCCGTAG).
21. hTRT mRNA was amplified using oligonucleotide primers LT5 and LT6 (19) for 31 cycles (94°C for 45 s, 60°C for 45 s, 72°C for 90 s). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA was amplified using primers K136 (CTCAGACACCAT-GGGGAAGGTGA) and K137 (ATGATCTTGAG-GCTGTTGTCATA) for 16 cycles (94°C for 45 s, 55°C for 45 s, 72°C for 90 s). hTR was amplified using primers F3b (TCTAACCCTAACTGAGAAGGGCG-TAG) and R3c (GTTTGCTCTAGAATGAACGGTG-GAAG) for 22 cycles (94°C for 45 s, 55°C for 45 s, 72°C for 90 s). TP1 mRNA was amplified using primers TP1.1 (TCAAGCCAAACCTGAATCTGAG) and TP1.2 (CCCGAGTGAATCTTTCTACGC) for 28 cycles (cycles same as for hTRT). Reaction products were resolved on an 8% polyacrylamide gel, stained with SYBR Green I (Molecular Probes, Eugene, OR).
22. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* 256, 1783 (1992).
23. A. Jacobo-Molina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 90, 6320 (1993); P. H. Patel *et al.*, *Biochemistry* 34, 5351 (1995).
24. Y. Xiong and T. H. Eickbush, *EMBO J.* 9, 3353 (1990); T. H. Eickbush, in *The Evolutionary Biology of Viruses*, S. S. Morse, Ed. (Raven, New York, 1994), pp. 121–157.
25. M. D. Powell *et al.*, *J. Biol. Chem.* 272, 13262 (1997).
26. M. Ricchetti and H. Buc, *Biochemistry* 35, 14970 (1996).
27. H. Biessmann *et al.*, *Cell* 61, 663 (1990); R. W. Levis, R. Ganesan, K. Houtchens, L. A. Tolar, F. Sheen, *ibid.* 75, 1 (1993); M. L. Pardue, O. N. Danilevskaya,
K. Lowenhaupt, F. Slot, K. L. Traverse, *Trends Genet.* 12, 48 (1996).
28. *h⁺/h⁻ leu1-32/leu1-32 ura4-D18/ura4-D18 ade6-M210/ade6-M216 his3-D1/his3-D1 trt1⁺/trt1⁻::his3⁺*.
29. C. Alfa, P. Fantes, J. Hyams, M. McLeod, E. Warbrick, *Experiments with Fission Yeast* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993).
30. N. Saitou and M. Nei, *Mol. Biol. Evol.* 4, 406 (1987). GenBank protein identification numbers of the sequences used in the phylogenetic analysis can be obtained from the authors. Alignment was analyzed using ClustalW 1.5 [J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* 22, 4673 (1994)] and PHYLIP 3.5 [J. Felsenstein, *Cladistics* 5, 164 (1989)].
31. We thank R. Adams, B. Lastelic, L. Tonkin, and F. Wu for expert technical assistance; C. Chapon, J. P. Cooper, R. Gutell, E. Jabri, and J. Sperger for discussions; R. Allshire and J. A. Wise for plasmids and yeast strains; C. Mattison and the L. Pillus lab for help with microscopy; and A. Sirimarco for manuscript preparation. An *S. pombe* cDNA library was provided by C. J. Norbury and B. Edgar. Supported by NIH grant GM28039 (T.R.C.).

Contrasting Genetic Influence of *CCR2* and *CCR5* Variants on HIV-1 Infection and Disease Progression

Michael W. Smith,* Michael Dean,* Mary Carrington,*
Cheryl Winkler, Gavin A. Huttley, Deborah A. Lomb,
James J. Goedert, Thomas R. O'Brien, Lisa P. Jacobson,
Richard Kaslow, Susan Buchbinder, Eric Vittinghoff,
David Vlahov, Keith Hoots, Margaret W. Hilgartner,
Hemophilia Growth and Development Study (HGDS),
Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia
Cohort Study (MHCS), San Francisco City Cohort (SFCC),
ALIVE Study, Stephen J. O'Brien†

The critical role of chemokine receptors (CCR5 and CXCR4) in human immunodeficiency virus–type 1 (HIV-1) infection and pathogenesis prompted a search for polymorphisms in other chemokine receptor genes that mediate HIV-1 disease progression. A mutation (*CCR2-64I*) within the first transmembrane region of the *CCR2* chemokine and HIV-1 receptor gene is described that occurred at an allele frequency of 10 to 15 percent among Caucasians and African Americans. Genetic association analysis of five acquired immunodeficiency syndrome (AIDS) cohorts (3003 patients) revealed that although *CCR2-64I* exerts no influence on the incidence of HIV-1 infection, HIV-1–infected individuals carrying the *CCR2-64I* allele progressed to AIDS 2 to 4 years later than individuals homozygous for the common allele. Because *CCR2-64I* occurs invariably on a *CCR5-+*–bearing chromosomal haplotype, the independent effects of *CCR5-Δ32* (which also delays AIDS onset) and *CCR2-64I* were determined. An estimated 38 to 45 percent of AIDS patients whose disease progresses rapidly (less than 3 years until onset of AIDS symptoms after HIV-1 exposure) can be attributed to their *CCR2-+/+* or *CCR5-+/+* genotype, whereas the survival of 28 to 29 percent of long-term survivors, who avoid AIDS for 16 years or more, can be explained by a mutant genotype for *CCR2* or *CCR5*.

The nexus of chemokine immunobiology and AIDS pathogenesis has revealed untapped avenues for resolving patterns of HIV-1 disease progression, for clarifying epidemiologic heterogeneity, and for design of therapies (*1–6*). Identification of the CC-chemokines, RANTES, MIP1α and MIP1β, as suppressor factors produced by CD8 cells that counter infection by certain HIV-1 strain infections (*7*) previewed the critical identification of two chemokine receptor molecules, CXCR4 (formerly named LESTR/fusin) and CCR5 (formerly CKR5), as cell surface coreceptors with CD4 for HIV-1 infection (*8–13*). Additional chemokine receptors CCR2 and CCR3 also have been implicated as HIV-1 coreceptors on certain cell types (*12–14*). HIV-1–infected patients harbor predominantly macrophage-tropic HIV-1 isolates during early stages of infection, but accumulate increasing amounts of T cell–tropic strains just before accelerated T cell depletion and progression to AIDS. The identification of "dual"-tropic HIV-1 strains over the course of infection suggests that such strains may represent an intermediate between macrophage- and T cell–tropic populations (*11–13, 15*). This tropic transition indicates that viral adaptation from CCR5 to CXCR4 receptor use may be a key step in progression to AIDS (*16*).

Telomerase catalysis: A phylogenetically conserved reverse transcriptase

Victoria Lundblad

This information is current as of February 2007.

Notes:

Commentary

Telomerase catalysis: A phylogenetically conserved reverse transcriptase

Victoria Lundblad

Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Replication of telomeres, the ends of eukaryotic chromosomes, is the responsibility of the enzyme telomerase. Since its discovery 13 years ago, research on this unusual DNA polymerase has revealed a series of surprises. The first of these was the realization that information within the enzyme itself determines the sequence of its product: a portion of a telomerase RNA subunit is the template that dictates the nucleotides added onto the telomere (1, 2). The interest in telomere replication has increased further during the past several years because of observations indicating that maintenance of telomere length by telomerase could provide the molecular basis for determining the lifespan of cells in culture (3).

The most recent insight has been the discovery that telomerase is a reverse transcriptase (4, 5), with catalysis provided by a protein subunit with striking similarities to conventional reverse transcriptases. The genes encoding the TERT (for telomerase reverse transcriptase; reviewed in ref. 6) proteins have been recovered from a diversity of species. Each of these proteins exhibits sequence features previously observed in reverse transcriptases, as well as a telomerase-specific T motif (ref. 7 and references therein). Mutation of key residues predicted to be critical for catalysis (by comparison to the reverse transcriptase active site) abolishes telomerase activity in yeast and humans (4, 5, 8), and expression of the human RNA and TERT subunits in an *in vitro* translation system is sufficient to reconstitute activity (8, 9). Although reliance on a templating RNA component had already suggested parallels between this enzyme and other RNA-dependent DNA polymerases, the demonstration that the telomerase catalytic subunit exhibited structural and enzymatic similarities to conventional reverse transcriptases provided direct mechanistic support for these comparisons. This result also has striking implications for reverse transcriptase evolution (10, 11), by demonstrating that such enzymes are not employed solely for replication of parasitic genetic elements but are also necessary for normal cellular proliferation. Furthermore, the cloning of the human TERT (hTERT) subunit permitted a direct test of the hypothesized role of telomere replication in the lifespan of normal human cells in culture: ectopic expression of hTERT in telomerase-negative human diploid fibroblasts restored enzyme activity and conferred an ability to proliferate well beyond the normal senescence point (12).

In this issue of the *Proceedings*, one aspect of telomerase research has come full circle, with the cloning of the TERT protein from the ciliate *Tetrahymena* (7, 13), the source of the first discovered telomerase activity. The ciliated protozoa have contributed greatly to our understanding of telomere biology because of an unusual feature of ciliate development. During the formation of a new macronucleus after mating, *de novo* addition of telomeres occurs on the ends of hundreds of thousands of newly formed minichromosomes (14). Thus, ciliates such as *Tetrahymena, Euplotes,* and *Oxytricha* have proven to be rich sources of the factors required for telomere

replication and maintenance. As a consequence, the first two telomerase-associated proteins, p80 and p95, were identified after purification of the *Tetrahymena* telomerase complex (15). On the basis of limited sequence similarities with other polymerases, p95 was proposed to contain the catalytic active site of this enzyme (15). However, p95 showed no homology to the emerging family of TERT proteins. This presented a potential puzzle, invoking the possibility of an alternative class of telomerase enzymes that utilized a different catalytic mechanism.

This possibility has now been laid to rest by two reports in this issue, from the Cech and Collins laboratories, showing that the *Tetrahymena* telomerase relies on a reverse transcriptase subunit for catalysis (7, 13). The gene encoding the *Tetrahymena* TERT protein was cloned by using a molecular approach, and the predicted protein displayed both the expected seven reverse transcriptase motifs and the T motif. Expression of this TERT protein and the RNA subunit in reticulocyte extracts was sufficient to reconstitute polymerization activity (although the high processivity observed with native enzyme was not achieved with this reconstituted core complex), and catalysis by the reconstituted enzyme was abolished by mutations similar to those previously tested for yeast and human telomerases (13). The *Tetrahymena* catalytic subunit protein also could be coimmunoprecipitated with p95, p80, and the RNA subunit, arguing that all four components are present in a single complex (13). Expression of the TERT mRNA was observed to increase dramatically after mating (7), consistent with the greatly increased requirements for telomere addition during macronuclear development.

Thus, it is now clear that the well characterized *Tetrahymena* enzyme has a catalytic subunit that shows both structural and evolutionary conservation with other telomerases. This result argues for the conceptually satisfying view that telomerase, regardless of its source, has a phylogenetically conserved core that minimally consists of the RNA component and a TERT protein. However, this conservation does not hold up as well when potential holoenzyme-associated proteins, isolated in several different systems, are compared. The telomerase reverse transcriptase subunit was discovered and characterized as a result of parallel biochemical and genetic endeavors in *Euplotes* and *Saccharomyces cerevisiae*; these efforts also led to the identification of telomerase-associated proteins in both species. Extensive purification of an active enzyme from *Euplotes aediculatus* resulted in a complex consisting of the RNA and catalytic protein subunits, as well as a copurifying 43-kDa protein; however, p80- and p95-like proteins were not detected in this complex (16). An alternative strategy for the identification of yeast telomerase components relied on a screen for mutants of yeast that exhibited an *in vivo* telomere replication defect (17, 18). This uncovered *EST2*, the yeast homolog of the telomerase reverse transcriptase gene, as well

---

The companions to this commentary are published on pages 8479–8484 and pages 8485–8490.
*To whom reprint requests should be addressed. e-mail: lundblad@bcm.tmc.edu.

as additional *EST* genes that, when mutated, exhibited a mutant phenotype identical to that displayed by mutations in the core enzyme complex. Est1, Est2, and Est3 proteins each coimmunoprecipitate with yeast telomerase activity (ref. 19; T. R. Hughes, R. G. Weilbaecher, and V.L., unpublished observations), suggesting that Est1 and Est3 could be components of the yeast holoenzyme. However, the yeast enzyme has not been purified sufficiently to be able to assess whether all three Est proteins are present in a single complex with the telomerase RNA. Strikingly, the telomerase-associated Est proteins show no similarity to any of the ciliate proteins. Furthermore, a search of the completely sequenced yeast genome has not revealed any homologs of the *Tetrahymena* p80 and p95 proteins.

So what is the explanation for the lack of convergence between these different sets of telomerase-associated proteins? Several hypotheses come to mind, which are not necessarily exclusive. The first possibility is that telomerase may be a large holoenzyme with a number of associated proteins, and efforts in these three organisms have succeeded in identifying only a partial subset. In support of this possibility, human, rat, and murine homologs of p80 have been isolated (20, 21), and the human p80 homolog has been shown to be in a complex with the hTERT subunit of telomerase (22), consistent with a similar demonstration for the equivalent *Tetrahymena* proteins (13). This finding excludes the possibility that p80 is a ciliate-specific telomerase protein and also raises the expectation that a similar mammalian p95 homolog may follow. This cross-species conservation does make the lack of a recognizable yeast version even more puzzling, but perhaps yeast homologs may not be readily identified on the basis of primary sequence. In fact, one proposed set of orthologs may be p95 and the yeast Est1 protein, as these two telomerase-associated proteins have a similar set of *in vitro* biochemical properties. Both proteins exhibit low-affinity, but sequence-specific, binding to single-strand telomeric DNA substrates (23, 24). In addition, both proteins interact, albeit nonspecifically, with RNA *in vitro* (23, 24). Such properties argue for roles in recognition of the telomeric DNA substrate and interaction with the telomerase RNA.

Alternatively, the diversity of telomerase-associated proteins may be a reflection of the differing requirements faced by telomerase in various biological situations, such that although the enzyme core may be conserved, at least a subset of the proteins that associate with the holoenzyme will be species-specific. One obvious species difference that may be mediated by components of the telomerase holoenzyme is the substantial variation in telomere length, ranging from 50 bp for some ciliates to ≥50 kb in one species of mouse. In addition, an enzyme that is responsible for *de novo* telomere addition as a part of chromosome healing might be expected to have differing cofactor requirements than an enzyme complex that is responsible for telomere length maintenance during vegetative growth. Even within a single species (*Euplotes crasses*), biochemical differences have been noted between telomerase isolated from vegetatively growing cells versus enzyme from mated cells (25).

Thus, although the basis for the differences in holoenzyme composition between different species is not yet understood, the results of the last year have shown that a phylogenetically conserved TERT protein is common to all telomerases. This has provided specific insight into the mechanism of telomerase catalysis, as well as establishing a sound foundation for a future detailed understanding of the composition of the telomerase holoenzyme. With four components of the most thoroughly studied telomerase now identified and available, *Tetrahymena* once again establishes itself as a system that will contribute important information about the biochemical activities of the components of this unusual DNA polymerase.

1. Greider, C. W. & Blackburn, E. H. (1989) *Nature (London)* **337**, 331–337.
2. Yu, G. L., Bradley, J. D., Attardi, L. D. & Blackburn, E. H. (1990) *Nature (London)* **344**, 126–132.
3. Autexier, C. & Greider, C. W. (1996) *Trends Biochem. Sci.* **21**, 387–391.
4. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T. R. (1997) *Science* **276**, 561–567.
5. Counter, C. M., Meyerson, M., Eaton, E. N. & Weinberg, R. A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 9202–9207.
6. Nugent, C. I. & Lundblad, V. (1998) *Genes Dev.* **12**, 1073–1085.
7. Bryan, T. M., Sperger, J. M., Chapman, K. B. & Cech, T. R. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8479–8484.
8. Weinrich, S. L., Pruzan, R., Ma, L., Ouellette, M., Tesmer, V. M., Holt, S. E., Bodnar, A. G., Lichsteiner, S., Kim, N. W., Trager, J. B., *et al.* (1997) *Nat. Genet.* **17**, 498–502.
9. Beattie, T., Zhou, W., Robinson, M. & Harrington, L. (1998) *Curr. Biol.* **8**, 177–180.
10. Eickbush, T. H. (1997) *Science* **277**, 911–912.
11. Nakamura, T. M. & Cech, T. R. (1998) *Cell* **92**, 587–590.
12. Bodnar A. G., Ouellette M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S. & Wright, W. E. (1998) *Science* **279**, 349–352.
13. Collins, K. & Gandhi, L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8485–8490.
14. Prescott, D. M. (1994) *Microbiol. Rev.* **58**, 233–267.
15. Collins, K., Kobayashi, R. & Greider, C. W. (1995) *Cell* **81**, 677–686.
16. Lingner, J. & Cech, T. R. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10712–10717.
17. Lundblad, V. & Szostak, J. W. (1989) *Cell* **57**, 633–643.
18. Lendvay, T. S., Morris, D. K., Sah, J., Balasubramanian, B. & Lundblad, V. (1996) *Genetics* **144**, 1399–1412.
19. Steiner, B. R., Hidaka, K. & Futcher, B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 2817–2821.
20. Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M. B., Arruda, I. & Robinson, M. O. (1997) *Science* **275**, 973–977.
21. Nakayama, J., Saito, M., Nakamura, H., Matsuura, A. & Ishikawa, F. (1997) *Cell* **88**, 875–884.
22. Harrington, L., Zhou, W., McPhail, T., Oulton, R., Yeung, D. S. K., Mar, V., Bass, M. B. & Robinson, M. O. (1997) *Genes Dev.* **11**, 3109–3115.
23. Virta-Pearlman, V., Morris, D. K. & Lundblad, V. (1996) *Genes Dev.* **10**, 3094–3104.
24. Gandhi, L. & Collins, K. (1998) *Genes Dev.* **12**, 721–733.
25. Bednenko, J., Melek, M., Greene, E. C. & Shippen, D. E. (1997) *EMBO J.* **16**, 2507–2518.

Phylogenetic Relationships of Reverse Transcriptase and RNase H Sequences and Aspects of Genome Structure in the Gypsy Group of Retrotransposons[1]

Mark S. Springer, * *and Roy J. Britten†*

*Department of Biology, University of California; and †Division of Biology, California Institute of Technology

The *gypsy* group of long-terminal-repeat retrotransposons contains elements having the same order of enzyme domains in the *pol* gene as do retroviruses. Elements in the *gypsy* group are now known from yeast, filamentous fungi, plants, insects, and echinoids. Reverse transcriptase and RNase H amino acid sequences from elements in the *gypsy* group—including the recently described *SURL* elements, *TED, Cft1,* and *Ulysses,*—were aligned and analyzed by using parsimony and bootstrapping methods, with plant caulimoviruses and/or retroviruses as outgroups. Clades supported at the 95% level after bootstrapping include (1) *17.6* with *297* and (2) all of the *SURL* elements together. Other likely relationships supported at lower bootstrap confidence intervals include (1) *SURL* elements with *mag,* (2) *17.6* and *297* with *TED,* and this collective group with *412* and *gypsy,* (3) *Tf1* with *Cft1,* (4) *IFG7* with *Del,* and (5) all of the retrotransposons in the *gypsy* group together, to the exclusion of *Ty3.* In contrast with an earlier analysis, our results place *mag* within the *gypsy* group rather than outside of a cluster that contains *gypsy* group retrotransposons and plant caulimoviruses. Several features of retrotransposon genomes provide further support for some of the aforementioned relationships. The union of *SURL* elements with *mag* is supported by the presence of two RNA binding sites in the nucleocapsid protein. Location of the tRNA primer binding site and the presence of a long open reading frame 3' to the *pol* gene support the *17.6-297-TED-412-gypsy* cluster.

Introduction

Retrotransposons containing long terminal repeats (LTRs) have now been identified in the genomes of a number of organisms and can be divided into two groups on the basis of both phylogenetic analysis of amino acid sequences and structural features of the genome (Xiong and Eickbush 1988, 1990; Doolittle et al. 1989). In the *copia* group, with representatives from *Drosophila* (*copia* and *1731*), yeast (*Ty1*), plants (*Tnt1, Ta1-3, Tst1, Wis,* and *Bis*), and *Physarum* (*Tp1*), the integrase gene is located between the protease and reverse transcriptase genes. In the *gypsy* group, with representatives from insects (*gypsy, 412, 17.6, 297, mag, micropia,* and *Ulysses*), yeast (*Ty3* and *Tf1*), filamentous fungi (*Cft1*), echinoids (*SURL* elements), and plants (*IFG7* and *Del*), the integrase gene is located 3' to the RNase H gene. The *gypsy* group of LTR retrotransposons is related to plant caulimoviruses and to retroviruses, on the basis of reverse transcriptase sequences (Xiong and Eickbush 1990).

---

1. Key words: retrotransposon, reverse transcriptase, RNase H.

Here, we examine phylogenetic relationships among members of the *gypsy* group by using amino acid sequences from the reverse transcriptase and RNase H proteins. Previous phylogenetic analyses of the *gypsy* group include Doolittle et al. (1989) and Xiong and Eickbush (1990). Xiong and Eickbush (1990) included 10 elements from the *gypsy* group in their analysis of reverse transcriptase sequences. Since that time, sequences for *SURL* elements, *TED*, *Tf1*, *Cft1*, and *Ulysses* have become available. We also evaluate the distribution and evolution of structural features in these retrotransposons in the light of amino acid–based phylogenies. Several structural features corroborate phylogenetic analysis on the basis of amino acid sequences.

Methods

Amino acid sequences and features of retrotransposons were obtained from GenBank and from references given in figure 1. Sequences of representative plant caulimoviruses were also obtained from GenBank. Delineation of boundaries for the reverse transcriptase protein correspond to that used by Xiong and Eickbush (1990). Delineation of RNase H sequence boundaries roughly corresponds to the region identified by McClure (1991). Multiple alignments were made by using CLUSTAL (Higgins and Sharp 1988), and adjustments were made by eye when conserved residues defined in Xiong and Eickbush (1990) and McClure (1991) were not aligned. Maximum parsimony and bootstrapping were performed by using PAUP, version 3.0s (Swofford 1991), with gaps counted as missing data. Plant caulimoviruses and/or retroviruses were used as outgroups. Each step on a parsimony tree corresponds to a single amino acid replacement. Because exact methods of finding minimum-length trees could not be used for the complete set of sequences, a heuristic approach using 100 replications with random input orders was employed. We also used a starting tree consistent with the tree given in Xiong and Eickbush (1990) as a baseline for searching for shorter trees. A distance matrix based on the aligned amino acid sequences was constructed by using the Kimura (1983, p. 175) option of the PROTDIST program on PHYLIP and was analyzed by using the neighbor-joining method (Saitou and Nei 1987).

Results
Alignments

Figure 1 shows a multiple alignment of amino acid sequences from the reverse transcriptase region. Overall, this alignment is similar to that of Xiong and Eickbush (1990), and most of the conserved blocks in their alignment are retained in the present alignment. Figure 2 shows an alignment of sequences from the RNase H region.

Phylogenetic Trees

Two minimum-length trees containing 2,219 amino acid replacements were found for the combined reverse transcriptase/RNase H sequences. One of these trees, rooted by using plant caulimoviruses, is shown in figure 3. On the second tree (not shown) the *Tf1-Cft1* group and *IFG7-Del* groups switch positions, and *micropia* is closer to *Ulysses* than to the *SURL-mag* group. Also shown on the tree in figure 3 are the consensus results of 500 bootstrap replications. Results summarized in figure 3 show (1) a likely sister-group relationship (86%) of *TED* (from the cabbage looper *Trichoplusia ni*) with *17.6* plus *297* (from *Drosophila*), (2) a likely sister-group relationship (73%) between the plant retrotransposons *IFG7* and *Del*, (3) a likely sister-group relationship (80%) between *SURL* elements and *mag*, (4) a likely sister-group rela-

```
17.6         LNQGIIRTSNSP--YNSPIWUUPKKQDA------SGKQKFRIVIDYR-KLNEITUGDRHPIPNMDEILGKLGRC-NYFTTIDLAKGFHQIEMDPESUSKTAFST
297          LNQGLIRESNSP--YNSPTWUVPKKPDA------SGANKYRVVIDYR-KLNEITIPDRYPIPNMDEILGKLGKC-QYFTTIDLAKGFHQIEMDEESISKTAFST
TED          LDQGIIRPSDSA-WSSPIWUVPKKIDA------SGKQKWRLVUDFR-KLNEKTIDDKYPIPNISDVLDKLGKC-QYFTTLDLASGFYQVEMDPQDISKTAFNV
412          IKDKIVEPSVSQ--YNSPLLLUVPKKSSP---NSDKKKWRLVIDYR-QINKKLLADKFPLPRIDDILDQLGRA-KYFSCLDLMSGFHQIELDEGSRDITSFST
gypsy (Dm)   LKDGIIRPSRSP--YNSPTWUVDKKGTD----AFGNPNKRLVIDFR-KLNEKTIPDRYPMPSIPMILANLGKA-KFFTTLDLKSGYHQIYLAEHDREKTSFSV
gypsy (Dv)   LDDGIIRPSRSP--YNSPTWUVDKKGTD----SYGNPKKRLVIDFR-KLNEKTIPDRYPMPRIPMILANLGKA-KYFTTLDLKSGYHQIYLAEHDREKTSFSV
SURL (Sp)    M--DVITRVDEPTDWVSSLVUVMKKNGQ----LRVCLDPR-DLNRAIKREHYQLPSRAEITAHFAGA-KVFSKLDASSGFWQIQLDDESSKLCTFIT
SURL (Tg)    L--DVITPVDEPTDWVSSLVUVMKKNGQ----LRVCLDPR-DLNRAIKREHYQLPSRAEITAHFAGA-KVFSKLDASSGFWQIQLDDESSKLCTFIT
SURL (Lv)    M--DVIKQVDEPTDWVSSLVUVKKKNGQ----LRVCLDPR-DLNRAIKREHYKLPSRAEITSQFAGA-KYFSKLDASSGFWQIQLDEDSSKLCTFIT
mag          LAAGVIKPVDHS-DWATPLVUVRKADGG----LRICADYKVTLNKVLAIDRFPVPKMEDLFSNLSGN-KFFTKLDLSQAVNQIVLSERSSEYTVINT
micropia     IRCNIVRPSCSP--FASPMLLVKKKNGT----DRLCVDFR-ELNSNTISDKYPLPLISDQIARLRGA-NYFTCLDMASGFHQIPIHPESVEYTAF--
Ulysses      LKLGIIEESDSP--WSNRTTVVMRPGKN----RFCLDAR-KLNSVTVKDAYPLPCIEGILSRSTRL-ILSLASTLSSRSGNRDGGEEQGVVGVVCT
IFG7         LEAGIIQPSQSS--FSDPVVLVHKKDGS----WCMCPDVR-ELNKLTIKDKFPIPVIDELLDELHGS-IYFTKLDLRSGVHQIRMKTEDIPKTTFRT
Del          LNKGFIRGSTSP--WGAHVLFDPKKDDS----KRMCIDY*-KLNSVTVKNKYPLPRIDDLFDQLNGA-*VFSKIDLRFRYHQLRIRA*DIPKTAFRT
Tf1          LKSGIIRESKAI--NACPVMFVPKKEGT----LRMVVDYK-PLNKYVKPNIYPLPLPLIEQLLAKIQGS-TIFTKLDLKSAVHLIRVRKGDEHKLAFRC
Cft1         LAKGWIRRSTSS--AGTPCMFVPKANGK----LRLVVQDYR-KLNEITIKNRYPLPNIEEAQDRLTGS-DWYKIDLRDAFVAIRMAEGEWKTAFRT
Ty3          LDNKFIVPSKSP--CSSPVVLVPKKDGT----FRLCVDYR-TLNKATISDPFFLPRIDNLLSRIGNA-QIFTTLDLHSGVHQIPMEPKDRYKTAFVT
COYMV        LQMKVIRPSESK--HRSTAFIVRSGTEIDPITGKEKKGKERMVFNYK-LLNENTESDQYSLPGINTIISKVGRS-KIVSKFDLKSGFWQVAMEEESVPWTAFLA
CERV         LELKVIKPSKST--HMSPAFLVENEAER------RRGKKRMVUNYK-AMNKATKGDAHNLPNKDELLTLVRGK-KIVSSFDCKSGLWQVLLDKESQLLTAFTC
FIGWORT      LDLGLIIPSKSQ--HMSPAFLVENEAER------RRGKKRMVUNYK-AINQATIGDSHNLPNMQELLTLLRGK-SIFSSFDCKSGFWQVLDEESQKLTAFTC
CMV          LDLKVIKPSKSP--HMAPAFLVNNEAEK------RRGKKRMVUNYK-AMNKATVGDAYNLPNKDELLTLIRGK-KIFSSFDCKSGFWQVLLDQESRPLTAFTC
HIV-I        EGKISKIGPENP--YNTPVFAIKKKDST------KVRKLVDFR-ELNKRTQDFWEVQLGI-PHPAGLKKK-KSVTVLDVGDAYFSVPLDKDFRKYTAFTI
HUER         LEKGHIEPSFSP--WNSPVFVIQKKSGK------WHTLTDLR-AVNAVIQPMGPLQPGL-PSPAMIPKD-WPLIIDLKDCFFTIPLAEQDCEKFAFTI
MOUSE IAP    LKLGHIDPSTSP--WNTPIFVIKKKSGK------WRLLHDLR-PINEQMNLFGPVQRGL-PVLSALPRG-WNLIIDIKDCFFSIPLCPRDRPRFAFTI
RSV          LQLGHIEPSLSC--WNTPVFVIRKASGS------VRLLHDLR-AVNAKLVPFGAVQQGA-PVLSALPRG-WPLMVLDLKDCFFSIPLAEQDREAFAFTL
BLV          LEAGVISPWDGP--GNNPVFPVRKPNGA------WRFVHDLR-ATNALTKPIPALSPGP-PDLTAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTL
MoMLV        LDQGILVPCQSP--WNTPLLPVKKPGTN------DYRPVQDLR-EVNKRVEDIHPTUPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFRFEW
HTLV-I       LEAGHIEPYTGP--GNNPVFPVKKANGT------WRFIHDLR-ATNSLTIDLSSSSPGP-PDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTV
```

```
17.6        KH------  GHYEVLRMPFGLKNAPATFQRCMNDILRPLLNKH----  -CLUYLDDIIUFSTSLDE-HLQSLGLUFEKLAKANLKLQLDKCEFLKQETTFLGHUL
297         KS------  GHYEVLRMPFGLRNAPATFQRCMNNILRPLLNKH----  -CLUYLDDIIIFSTSLTE-HLNSIQLUFTKLADANLKLQLDKCEFLKKEANFLGHIV
TED         EH------  GHFEFLRMPMGLKNSPSTFQRVMDNVLRGLQNNI----  -CLUYLDDIIVYSTSLQE-HLENLERVFQRLRESNFKIQMDKSEFLKLETAYLGHII
412         SN------  GSYRFTRLPFGLKIAPNSFQRMMTIAFSGIEPSQ----  -AFLYMDDLIVIGCSEKH-MLKNLTEVFGKCREYNLKLHPEKCSFFMHEUTFLGHKC
gypsy (Dm)  NG------  GKYEFCRLPFGLRNASSIFQRALDDULREQIGKI----  -CYUVDDUIIFSENESD-HURHIDTVLKCLIDANMRUSQEKTRFFKESUEYLGFIV
gypsy (Dv)  SS------  GKYEFCRLPFGLRNASSIFQRAIDDILREHIGKI----  -CFUVDDUIIFSKNETE-HLQHINIVLKCLIDANMRUGPEKTRFFKESIEFLGFIV
SURL (Sp)   PY------  GRYKFLRLPFGICSAPEUVPKIUHQLFAHIPG------  -UNTMMDDVIVARTTQQE-RDDRLRKVLSIVRKMNLKLNKDKCEFNVKKLTFIGDLI
SURL (Tg)   PY------  GRYKFLRLPFGICSAPEUVHKIUHQMFAHIPG------  -UNTMMDDVIVHGTTQQE-HDNRLREVLSIARRMNLKLNKDKCEFSVKKLTFIGDLI
SURL (Lv)   PY------  GRYRFLRLPFGICSAPEUFHKIIHHLFUDIPG------  -UNTMMDDIVUHGSTQEE-HDDRLRKVLDIAQKSNLKLNRDKCEFNVNQMTFIGDLI
mag         HR------  GLFKYSRLUVGLASSPGIFQKLMUMFKNUPN------   -UUFVDDILIRNQDLDS-HLKSIKEVLDILERYGLKIKRSKCEFNVTEURYLGFII
micropia              --VPDGLKNAPSUFQRTUINALGDLANSF-----       -UIUVMDDIMUUSPTKEL-ALERLKTVLNULTKAGFTFNLAKCSFLKTTUQYLGYEU
Ulysses     RR------  PLYQFRHMPFGLCNAAQHFEA-HDKUIPANLRSN----  -UFUYLDDLIISADFPT-HLKYLELUAECLRNANLTIGMAKSKFLFRNLNYLGFIQ
IFG7        HE------  GHYEFFUMPFGLTNTPSTFQGLMNSIFKPFLRKF----  -ULUFFDDILIYNKSWKD-HUEHUDRVLQLLEEKKLYAKRSKCEFULQEUEYLGHIV
Del         RY------  GHYEFLUMPFGLTNUPTAFMNLMNRUFREYLDKF----  -IUUFUDYVLIYSRTQKD-HEHHLRISLQLLRNNQLYAKLSKCEFWMEKUKFLGHUV
Tf1         PR------  GUFEYLUMPYGISTAPAHFQYFINTILGEAKESH----  -UUCYMDDILIHSKSESE-HUKHVKDVLQKLKNANLIINQAKCEFHQSQUKFIGYHI
Cft1        RY------  GLYEFLUMPMGLTNAPASCQDLUNETLRDLLDUC----  -UUAYMDDILUYTKGSLQEHTKQUQDUFERLTKSGFKTAPEKCEFHKKEUKFLGFII
Ty3         PS------  GKYEYTUMPFGLUNAPSTFARVMADTFRDLRF----    -UNVYLDDILIFSESPEE-HWKHLDTVLERLKNEMLIUKKKKCKFASEETEFLGYSI
COYMV       GN------  KLYEWLUMPFGLKNAPAIFQR-KMDNUFKGTEKF----  -IAUVIDDILUFSETAEQ-HSQHLYTMLQLCKENGLILSPTKMKIGTPEIDFLGASL
CERV        PQ------  GHYQWNUUPFGLKQAPSIFPKTYANSHSNQYSKY----  -CCUVUDDILUFSNTGRKEHYIHULNILRACEKLGIILSKKKAQLFKEKINFLGLEI
FIGWORT     PQ------  GHFQWKUUPFGLKQAPSIFQR-HMQTALNGADKF----  -CMUVUDDIIUFSNSELD-HYNHUVAVLKIUEKYGIILSKKANLFKEKINFLGLEI
CMV         PQ------  GHYEWNUUPFGLKQAPSIFQR-HMDEAFRUFRKF----  -CCUVUDDIUFSNNEED-HLLHUAMILQKCNQHGIILSKKAQLFKKKINFLGLEI
HIV-I       PSINNETPGIRYQVNVLPQGGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQK-EPPFLWMGYEL
HUER        PAIMNKEPATRFQWKULPQGMLNSPTICQTFUGRALQPUREKFSDCYIHYIDDILCAAETKDK-LIDCYTFLQAEUANAGLAIASDKIQT-STPFHYLGMQI
MOUSE IAP   PSINSDEPDMRYQWKULPQGMSNSPTMCQLYVQEALLPUREQFPSLILLLYMDDILCHKELTM-LQKAYPFLLKTLSQWGLQIATEKUQI-SDTGQFLGSVU
RSV         PSUNNQAPARRFQWKULPQGMTCSPTICQLUVGQULEPLRLKHPSLCMLHYMDDLLLAASSHDG-LEAAGEEVISTLERAGFTISPDKVQR-EPGUQYLGYKL
BLV         PSPGGLQPHRRFAWRULPQGFINSPALFERALQEPLRQVSAAFSQSLLVSVMDDILYASPTEEQ-RSQCYQALAARLRDLGFQUASEKTSQTPSPUPFLGQMU
MOMLV       RDPE-MGISGQLTHTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQVUDDLLAATSELD-CQQGTRALLQTLGNLGYRASAKKAQICQKQUKYLGYLL
HTLV-I      PQQCNYGPGTRYAWRULPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHAD-LQLLSEATMASLISHGLPUSENKTQQTPGTIKFLGQII
```

FIG. 1.—Alignment of amino acid sequences from the reverse transcriptase region for the *gypsy* group of retrotransposons, several plant caulimoviruses, and several retroviruses. Abbreviations for plant caulimoviruses and retroviruses are as follows: COYMV = *Commelina* yellow mottle virus; CERV = carnation etched-ring virus; FIGWORT = figwort mosaic virus; CMV = cauliflower mosaic virus; HIV-I = human immunodeficiency virus I; HUER = human endogenous retrovirus K; MOUSE IAP = mouse intracisternal A particle; RSV = Rous sarcoma virus; BLV = bovine leukemia virus; MOMLV = Moloney murine leukemia virus; and HTLV-I = human T-cell leukemia virus I. An asterisk (*) denotes a stop codon. References for the *gypsy* retrotransposons are as follows: *17.6* (Saigo et al. 1984), *297* (Inouye et al. 1986), *412* (Yuki et al. 1986), *gypsy* (Marlor et al. 1986; Mizrokhi and Mazo 1991), *micropia* (Lankeneu et al. 1988), and *Ulysses* (Scheinker et al. 1990) are from *Drosophila*; TED (Friesen and Nissen 1990) is from the cabbage looper (*Trichoplusia ni*); SURL elements (Springer et al. 1991) are from echinoids; *mag* (Michaille et al. 1990) is from the silk moth (*Bombyx mori*); *IFG7* (Kossack 1989) is from *Pinus radiata*; *Del* (Sentry and Smyth 1989; Smyth et al. 1989) is from the lily (*Lilium henryi*); *Tf1* (Levin et al. 1990) is from fission yeast (*Schizosaccharomyces pombe*); *Cf11* (McHale et al. 1992) is from *Cladosporium fulvum*; and *Ty3* (Hansen et al. 1988) is from *Saccharomyces cerevisiae*.

```
17.6       DFTKKFTLTTDASDVALGAVLSQDGH--------PLSYISRTLNEHEINYSTIEKELLAIVHATKTFRHYLLGR
297        DFEKKFVLTTDASNLALGAVLSQNGH--------PISFISRTLNDHELNYSAIEKELLAIVHATKTFRHYLLGR
TED        DFTREFNLTTDASNFAIGAVLSQGPIGSDK-----PVCVASRTLNESELNYSTIEKELLAIVHATKVFRPVLFGR
412        DFSKEFCITTDASKQACGAVLTQNHNGHQL-----PVAYASRAFTKGESNKSTTEQELAAIHUAIIHFAPYIYGK
gypsy (Dm) DFKKPFDLTTDASASGIGAVLSQEGR--------PITMISRTLKQPEQNYATNERELLAIVHALGKLQNFLYGS
gypsy (Dv) NFQKPFDLTTDASASGIGAVLSQGNR--------PITHISRALKQAEQNYATNERELLAIVHALGRLQNFLYGS
SURL (Sp)  DCKKLTKLSADASKDGIGAVLLQQVDQDHV-----PIAVASRSMTDAETRYAQIEKELLAITYACERFHQYIYGQ
SURL (Tg)  DCNKPTKLSADASKNGIGAVLLQQHDENHV-----PIAVASRSMTDAETRYAQIEKELLAITYACERFHQYIYGQ
SURL (Lv)  DCTKPTKISADASKNGLGAVLLQQHEQNHH-----PIAVASRAMTDAETRYAQIEKELLAITYGCEKFHQYIYGQ
mag        DHSLESVLTVDASARGLGAVLAQRGPGCQER----VVAYASRALTTHELHVSQIHKEALAIVFAVEKFHQYLYGR
micropia   DPQYPIELHTDASACGYGAILLHRIESKPH-----VIEVFSKTTTSVESRVHSYELETLAVVKAVKHFRHYLIGR
Ulysses    DFRRPFFIQCDASHVGVGAVLFQLDDEQQER----PIAFFSAKLNKHQINVSUTEKECLAAKLAIHRFRPYVEMM
IFG7       DFMKTFIVECDASGNGIGAVLMQDEI--------PIAFEGHPIRGKFLHKALYEKEMLAILHALKKHRPYLMGR
Del        ISG*PFUVYTDASLAGLEGVLMQDGR--------VVAYASRQLKVHENMYPTHDLELAVVIFILKLHRHYLYGE
Tf1        DFSKKILLETDASDVAVGAVLSQKHDDDKYY----PVGVYSAKMSKAQLNYSVSDKEMLAIIKSLKHHRHYLEST
Cft1       DGSKEVHIETDASDMAIGACLTQTHDGKRH-----PVAVYSRKMTTAEQNYDIHDKELLAIVAAMQHHRUVYUEGP
Ty3        NNKANYRLTTDASKDGIGAVLEEVDNKNKLVG---VVGYFSKSLESAQKNYPAGELELLGIIKALHHFRYMLHGK
COYMV      PKDSFIIIETDGCHTGHGAVCKHKMSKHDPRSTERICAYASGSFNPIKS---TIDAEIQAAIHGLDKFKIYYLDK
CERV       EPNDKLVIETDASEEFHGGILKAIHN-----SHEVICRYASGSFKAAERNYHSNEKELLAVIRVIKKFSIYL·TPS
FIGWORT    KPEDHLIIETDASDSFHGGVLKARALD----GVELICRYSSGSFKQAEKNYHSNDKELLAVKQVITKFSAYLTPV
CMV        LPEEKLIIETDASDDVHGGMLKAIKIN-EGTNTELICRYASGSFKAAEKNYHSNDKETLAVINTIKKFSIYLTPV
```

```
17.6       --HFEISSDHQPLSHLVRMK-----DPNS-KLTRH---RVKLSEFDFDI--KYIKGKENCVADALSRIKLEETY
297        --QFLIASDHQPLRHLHNLK-----EPGA-KLERH---RVALSEYQFKI--DYIKGKENSVADALSRIKIEENH
TED        --KFKILTDHKPLQHMMNLK-----DPNS-RMTRH---RLRLSEYDFSV--VYKKGKSNTNADALSRVEIHTTE
412        --HFTUKTDHRPLTYLFSMV-----NPSS-KLTRI---RLELEEYNFTV--EVLKGKDNHVADALSRITIKELK
gypsy (Dm) -REINIFTDHQPLTFAVADR-----NTNA-KIKRH---KSYIDQHNAKV--FYKPGKENFVADALSRQNLNALQ
gypsy (Dv) -REINIFTDHQPLTFAVSDK-----NTNS-KIKRH---KSYIDQHNAKM--FYKPGKENLVADALSRQNINALE
SURL (Sp)  --QVEVETDHKPHIPLFVKSLG---DCPL-RIQRL---LIRVQRYDLKV--VVTPGKVMYTADTLSRAVDPKAE
SURL (Tg)  --QVEVETDHKPLIPLFVKSLG---DCPL-RIQRL---LIRVQRYDLKV--MYTPGKVMYTADTLSRAVDPKAD
SURL (Lv)  --KIEVETDHKPLIPLFVKSLA---DCPL-RIQRL---LIRVQRYDLKU--SYTPGKVMFTADTLSRAVDPKAE
mag        --KFILRTDHKPLVSIFGPNIGIPSAAAS·RLQRH---AIKLSAYDFEI--EYVRT-DKNVADALSRLIESQKN
micropia   --EFUVVYTDCNSLKASRTKI-----DLTP-RVHRH---HAYLQSFNFEI--QYREGKRMAHVDFLSRNPLSPEH
Ulysses    --PFTVITDHASLQHLMSLK-----DLSG-RLARH---SLELQAFPFSM--QYRKGADNVCRHIVRSVEEVELT
IFG7       --HFNVKTDHDSLKYFLEQR-----LSSE-E*QKH---VTKMLGYDFEI--IYKKGKKNVVANALSRKDEDVEA
Del        --DFELVCDHKSLKYISTQK-----DLML-RQR*H---IEVLKDFDFSI--FYHPGKANVVADALSRKSQISHL
Tf1        IEPFKILTDHRNLIRITNGE----SEPENKRLARH---QLFLQDFNFEI--NVRPGSANHIADALSRIVDETEP
Cft1       -PKLTILSDHKNLTYFTTTK----ELTA-RQARH---SELLGQYKFEI--KYTPGTENGPADALSRRSDYMEG
Ty3        --HFTLRTDHISLLSLQNKN-----EPAR-RVQRH---LDDLATYDFTL--EYLAGPKNVVADAISRAIYTITP
COYMV      -KELIIRSDCEAIIKFYN---KTNENKP--SRVRHLTFSDFLTGLGITVTFEHIDGKHNGLADALSRMINFIVE
CERV       --RFLIRTDMKNFTHFVNINLKGDRKQG--RLVRH---QMHLSQYDFDV--EHIAGTKNVFADFLQENTLTNYV
FIGWORT    --RFTURTDMKNFTYFLRINLKGDSKQG--RLVRH---QNHFSKYQFDV--EHLEGUKNVLADCLTRDFNA---
CMV        --HFLIRTDNTHFKSFVNLNYKGDSKLG--RNIRH---QAHLSHYSFDV--EHIKGTDNHFADFLSREFNKUNS
```

FIG. 2.—Alignment of amino acid sequences from the RNase H region for the *gypsy* group of retro-transposons and several plant caulimoviruses. Abbreviations are given in fig. 1. An asterisk (*) denotes a stop codon.

tionship (81%) between *Tf1* (from fission yeast) and *Cft1* (from the fungal tomato pathogen *Cladosporium fulvum*), and (5) a possible clade (62%) containing *17.6, 297, TED, 412,* and *gypsy.* In addition, *Ty3* is an outgroup to all other retrotransposons in the *gypsy* group on 70% of the bootstrap trees. *Ulysses* and *micropia* group with *SURL* elements and *mag* on both minimum-length trees, but this association does not hold up after bootstrapping. Likewise, the minimum-length tree shown in figure 3 supports a clade containing all of the retrotransposons that occur in metazoans, but this branch does not occur on the second minimum-length tree, nor is it supported by bootstrapping. In contrast to the tree in figure 3, the shortest tree consistent with that of Xiong and Eickbush (1990) is 30 steps longer, at 2,249 steps.

When we converted our sequence alignments to distances by using the Kimura option of PROTDIST (PHYLIP, version 3.5; Felsenstein 1993) and then employed the neighbor-joining method, the resulting tree (not shown) showed some differences from the minimum-length trees, but all of the branches that are supported at the 50% level in figure 3 are also supported on the neighbor-joining tree.

Minimum-length trees (not shown) based on reverse transcriptase versus RNase H sequences exhibit several conflicts; for example, *SURL* elements cluster with *mag* on reverse transcriptase trees but cluster with the two *gypsy* elements on RNase H trees. However, all of the conflicts involve branches that are not supported after boots-
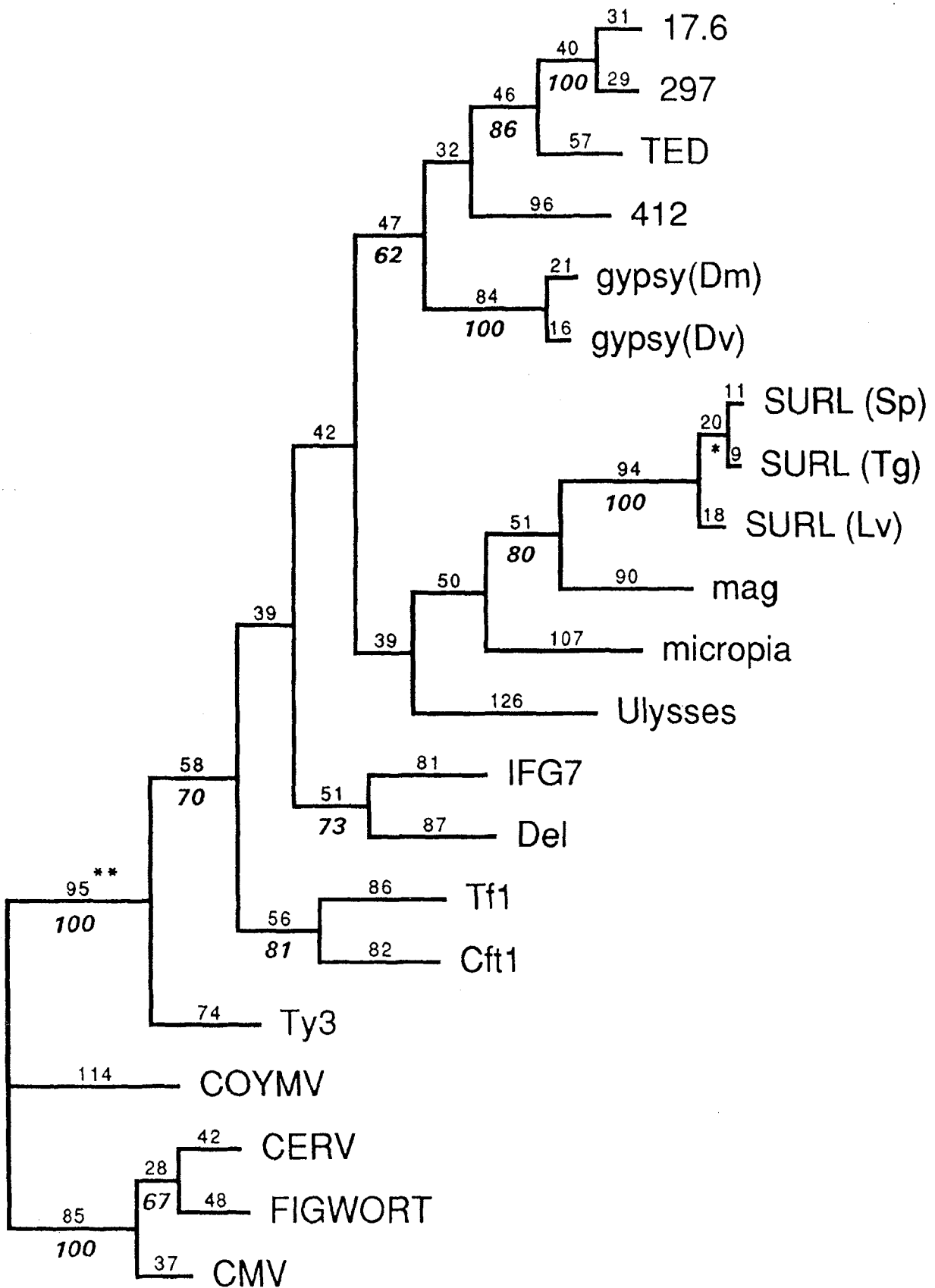
FIG. 3.—One of two minimum-length trees at 2,219 replacements. Numbers above the line are the number of amino acid replacements, and numbers below the line are the percentages, from 500 bootstrap trials, that support the clade. One asterisk (*) denotes a value of 99%, and two asterisks (**) denote the branch of this tree that the root would be on if the reverse transcriptases of the seven retroviruses were used for rooting.

trapping. Bootstrapping the reverse transcriptases and RNase H sequences, respectively, provides support for the following: *17.6* plus *297* (94% and 97%), and for this group with *TED* (85% and 64%); the two *gypsy* elements together (100% and 100%); the three *SURL* elements together (100% and 100%) with *SURL* (Sp) and *SURL* (Tg) as nearest neighbors (96% and 90%); and *Tf1* plus *Cft1* (60% and 73%). In addition, bootstrapping reverse transcriptase sequences provides support for *SURL* elements with *mag* (60%), *IFG7* plus *Del* (71%), and all of the retrotransposons together, except *Ty3* (52%).

Features of *gypsy*-like Elements

Table 1 summarizes the distribution of seven different features of retrotransposons in the *gypsy* group. The phylogenetic significance of these features is discussed below.

Discussion

Xiong and Eickbush (1988, 1990) previously examined relationships among retroid elements, including retrotransposons in the *gypsy* group, on the basis of reverse transcriptase sequences. One of the differences on the Xiong and Eickbush (1990) tree is that *mag* is outside of a cluster containing other retrotransposons in the *gypsy* group as well as plant caulimoviruses. To test this hypothesis with our data, it was necessary to include retroviruses as an outgroup to the collective group. We limited this analysis to reverse transcriptase sequences because of the difficulty in aligning RNase H sequences. Retroviruses clearly root the tree (not shown) such that the plant-caulimovirus and retrotransposon groups (including *mag*) are each monophyletic.

Two other differences on the Xiong and Eickbush (1990) tree are as follows: (1) *Ty3* is not peripheral to other *gypsy* retrotransposons but occupies a position close to *IFG7* and *Del,* and (2) *412* is the most peripheral member of the *gypsy* cluster, except *mag.* Whether we (1) use parsimony or neighbor-joining methods, (2) include RNase H and reverse transcriptase or just reverse transcriptase sequences, or (3) restrict our analysis to the reverse transcriptase sequences available to Xiong and Eickbush (1990), *Ty3* occupies the most peripheral position among retrotransposons in the *gypsy* group, and *412* clusters with the insect elements *gypsy, 297, 17.6,* and *TED.*

The overall congruence between reverse transcriptase and RNase H bootstrap trees indicates that a similar phylogenetic signal is present in both, although, when taken separately, each of these proteins provides less resolution than they do in combination with each other. One of the implications of the overall congruence between bootstrap trees is that reverse transcriptase and RNase H have similar evolutionary histories without any interelement recombination that might cause striking differences.

If *Ty3* is taken as an outgroup to all of the other retrotransposons, then the implied primitive character states for the characters in table 1 are +1 ribosomal frame-shifting, one RNA binding site, tRNA methionine, a +2 location of the tRNA primer binding site (PBS), and lack of a long open reading frame (ORF) 3' to the *pol* gene. On the basis of these designations of primitive character states, several of the aspects of genome structure given in table 1 offer additional support for some of the branches on the tree in figure 3. First, *17.6, 297,* and *TED* are united by the putative shared derived character of tRNA serine, although a putative tRNA serine also occurs in *Cft1* (McHale et al. 1992). Second, *17.6, 297, TED, 412,* and *gypsy* share a number of putative derived characters, including a long ORF 3' to the *pol* gene, a 1-bp overlap of the 5' LTR and the tRNA PBS, and −1 frameshifting of the *pol* gene relative to the *gag* gene, as well as the absence of RNA binding sites in the nucleocapsid protein. While two of these derived characters have evolved elsewhere on the tree (i.e., −1

Table 1
Select Features of Retrotransposons Related to the *gypsy* Element of *Drosophila*

FEATURE	RETROTRANSPOSON[a]													
	17.6	297	TED	412	gypsy[b]	SURL[c]	Mag	micropia	Ulysses	IFG7	Del	Tf1	Cft1	Ty3
1. Ribosomal frameshifting	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	No	Yes	Yes
No. of nucleotides[d]	−1	−1	−1	−1	−1	0	−1		+1				−1	+1
2. No. of RNA binding sites	0	0	0	0	0	2	2	2	0	1	1	0 (?)	1	1
3. tRNA PBS	Serine	Serine	Serine	Arginine	Lysine	Methionine	Arginine	Leucine	Lysine	Methionine	Methionine	?	Serine (?)	Methionine
4. Location of tRNA PBS[e]	−1	−1	−1	−1	−1	0	+1	0	+3	+2	+3	?	0	+2
5. Envelope (?) ORF[f]	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
6. LTR length (in bp)	512	415	273	481–571	479–482	254	77	474–504	2,136	333	2,406–2,415	358	427	340
7. Element length (in bp)	7,439	6,995	7,510	7,400	7,469	5,266	4,564	5,500	10,653	5,944	9,345	4,941	6,968	5,428

[a] Boxes enclose clades of fig. 3.

[b] Dm strain.

[c] Tg strain.

[d] No. of nucleotides that must "read" to change from the frame reading "xxx" to that reading "yyy."

[e] Relative to the 5′ LTR. Minus (−) values are base pair of overlap; plus (+) values are the distance between.

[f] A number of retrotransposons contain long (~1,500-bp) ORFs located between the *pol* gene and the 3′ LTR. In retroviruses, the ORF in this region codes for proteins that are essential in the extracellular stage of the life cycle; however, there is as yet no direct evidence that the similarly placed ORF that occurs in some retrotransposons codes for homologous proteins.

frameshifting also occurs in *Cft1* and *mag,* and RNA binding sites are absent in *Ulysses*), the presence of a long ORF 3' to *pol* and a −1 location of the tRNA PBS are unique to this subset of the *gypsy* group. Third, the putative relationship between *mag, SURL* elements, and possibly *micropia* is potentially strengthened by the exclusive occurrence of two RNA binding sites in the nucleocapsid protein in all of these elements. Most retroviruses also possess two RNA binding sites, but in the somewhat more closely related plant caulimoviruses there is only a single site. Further support for the alliance between *mag* and *SURL* elements comes from the observation that the number of amino acids separating the two RNA binding sites is identical in these elements. *Micropia,* in turn, has 14 additional amino acids that separate the first and second RNA binding sites. The plant elements *Del* and *IFG7* share a number of features, such as a single RNA binding site, a single ORF containing the *gag* and *pol* genes, and a tRNA methionine PBS, but these features appear primitive on the basis of their occurrence in *Ty3.*

The long LTRs in *Ulysses* and *Del* appear homoplastic on the basis of other evidence discussed above, whereas the short LTRs in *mag* are unique to this element. Element length ranges from 4,564 bp in *mag* to 10,653 bp in *Ulysses* and reflects the differences in LTR length. Among other elements, most of the variation results not from differences in LTR length but rather from the additional ORF 3' to the *pol* gene.

It is interesting that, for the tree in figure 3, all of the animal retrotransposons occur on one branch, whereas the two plant elements occur on a second branch. The separate clusters of plant and animal retrotransposons suggest that the host phylogeny imposes a distinct signature on the phylogeny of the retrotransposons; Flavell (1992) previously noted predominantly plant and animal groups for the *copia* group of retrotransposons as well. Flavell (1992) has also characterized the *copia* group as lacking ribosomal frameshifting, whereas in the *gypsy* group the *gag* and *pol* genes are always overlapping. However, the presence or absence of overlapping *gag* and *pol* genes is shown here to exhibit more variation in the *gypsy* group than was previously recognized.

In conclusion, our understanding of the phylogeny of the *gypsy* group of retrotransposons is enhanced by considering not only amino acid sequences but also genetic features of these elements. Some features (e.g., long 3' ORF) show little or no homoplasy, whereas others (e.g., type of tRNA PBS) are labile and show much more homoplasy.

Acknowledgment

LITERATURE CITED

DOOLITTLE, R. F., D.-F. FENG, M. S. JOHNSON, and M. A. MCCLURE. 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
FELSENSTEIN, J. 1993. PHYLIP manual, version 3.5. University of Washington, Seattle.
FLAVELL, A. J. 1992. *Ty1-copia* group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica **86**:203–214.
FRIESEN, P. D., and M. S. NISSEN. 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell. Biol. **10**:3067–3077.
HANSEN, L. J., D. L. CHALKER, and S. B. SANDMEYER. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol. Cell. Biol. **8**:5245–5256.

HIGGINS, D. G., and P. M. SHARP. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.

INOUYE, S., S. YUKI, and K. SAIGO. 1986. Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. Eur. J. Biochem. **154**:417–425.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KOSSACK, D. 1989. The IFG copia-like element: characterization of a transposable element present in high copy number in *Pinus* and a history of the pines using IFG as a marker. University Microfilms, University of Michigan, Ann Arbor.

LANKENEU, D.-H., P. HUIJSER, E. JANSEN, K. MIEDEMA, and W. HENNIG. 1988. Micropia: a retrotransposon of *Drosophila* combining structural features of DNA viruses, retroviruses and non-viral transposable elements. J. Mol. Biol. **204**:233–246.

LEVIN, H. L., D. C. WEAVER, and J. D. BOEKE. 1990. Two related families of retrotransposons from *Schizosaccharomyces pombe*. Mol. Cell. Biol. **10**:6791–6798.

McCLURE, M. A. 1991. Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol. Biol. Evol. **8**:835–856.

McHALE, M. T., I. N. ROBERTS, S. M. NOBLE, C. BEAUMONT, M. P. WHITEHEAD, D. SETH, and R. P. OLIVER. 1992. C f T-I: an LTR retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato. Mol. Gen. Genet. **233**:337–347.

MARLOR, R. L., S. M. PARKHURST, and V. G. CORCES. 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6**:1129–1134.

MICHAILLE, J.-J., S. MATHAVAN, J. GAILARD, and A. GAREL. 1990. The complete sequence of mag, a new retrotransposon in *Bombyx mori*. Nucleic Acids Res. **18**:674.

MIZROKHI, L. J., and A. M. MAZO. 1991. Cloning and analysis of the mobile element *gypsy* from *D. virilis*. Nucleic Acids Res. **19**:913–916.

SAIGO, K., W. KUGIMIYA, Y. MATSUO, S. INOUYE, K. YOSHIOKA, and S. YUKI. 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature **312**:659–661.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCHEINKER, V. SH., E. R. LOZOVSKAYA, J. G. BISHOP, and M. B. EVGEN'EV. 1990. A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *Drosophila virilis*. Proc. Natl. Acad. Sci. USA **87**:9615–9619.

SENTRY, J. W., and D. R. SMYTH. 1989. An element with long terminal repeats and its variant arrangements in the genome of *Lilium henryi*. Mol. Gen. Genet. **215**:349–354.

SMYTH, D. R., P. KALITSIS, J. L. JOSEPH, and J. W. SENTRY. 1989. Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of *Drosophila*. Proc. Natl. Acad. Sci. USA **86**:5015–5019.

SPRINGER, M. S., E. H. DAVIDSON, and R. J. BRITTEN. 1991. Retroviral-like element in a marine invertebrate. Proc. Natl. Acad. Sci. USA **88**:8401–8404.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.0s. Illinois Natural History Survey, Champaign.

XIONG, Y., and T. H. EICKBUSH. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5**:675–690.

XIONG, Y., and T. H. EICKBUSH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9**:3353–3362.

YUKI, S., S. INOUYE, S. ISHIMARU, and K. SAIGO. 1986. Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. Eur. J. Biochem. **158**:403–410.
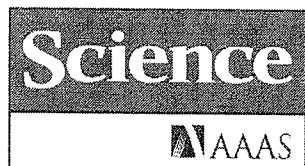
The following resources related to this article are available online at
www.sciencemag.org (this information is current as of February 22, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/276/5312/561

This article **cites 20 articles**, 16 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/276/5312/561#otherarticles

This article has been **cited by** 98 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/276/5312/561#otherarticles

This article appears in the following **subject collections**:
Molecular Biology
http://www.sciencemag.org/cgi/collection/molec_biol

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/help/about/permissions.dtl

Reverse Transcriptase Motifs in the Catalytic Subunit of Telomerase

Joachim Lingner,* Timothy R. Hughes, Andrej Shevchenko,
Matthias Mann,† Victoria Lundblad,† Thomas R. Cech†

Telomerase is a ribonucleoprotein enzyme essential for the replication of chromosome termini in most eukaryotes. Telomerase RNA components have been identified from many organisms, but no protein component has been demonstrated to catalyze telomeric DNA extension. Telomerase was purified from *Euplotes aediculatus*, a ciliated protozoan, and one of its proteins was partially sequenced by nanoelectrospray tandem mass spectrometry. Cloning and sequence analysis of the corresponding gene revealed that this 123-kilodalton protein (p123) contains reverse transcriptase motifs. A yeast (*Saccharomyces cerevisiae*) homolog was found and subsequently identified as *EST2* (ever shorter telomeres), deletion of which had independently been shown to produce telomere defects. Introduction of single amino acid substitutions within the reverse transcriptase motifs of Est2 protein led to telomere shortening and senescence in yeast, indicating that these motifs are important for catalysis of telomere elongation in vivo. In vitro telomeric DNA extension occurred with extracts from wild-type yeast but not from *est2* mutants or mutants deficient in telomerase RNA. Thus, the reverse transcriptase protein fold, previously known to be involved in retroviral replication and retrotransposition, is essential for normal chromosome telomere replication in diverse eukaryotes.

Replication of chromosome ends, or telomeres, requires specialized factors that are not essential for replication of internal chromosome sequences. Conventional DNA polymerases cannot fully replicate blunt-ended DNA molecules (*1*) or eukaryotic chromosomes (*2*), which contain 3'-terminal extensions. The key to end replication is telomerase, a ribonucleoprotein (RNP) enzyme that synthesizes the telomeric DNA repeats (*3*). The template for telomeric repeat synthesis is provided by the RNA subunit, which has been identified, cloned, and sequenced in ciliated protozoa (*4, 5*), yeast (*6, 7*), and mammals (*8*).

A telomerase RNP was first purified from *Tetrahymena* (*9*). Two protein components, p80 and p95, were specifically associated with the RNA subunit. Human, mouse, and rat homologs of *Tetrahymena* p80 have since been identified and found to be associated with telomerase (*10*). Although this evolutionary conservation suggests that p80 and p95 have important roles

in telomere replication, their specific functions remain unclear. Neither protein has been reported to be essential for telomere synthesis, and neither has significant simi-

larity to known polymerases or reverse transcriptases (*11*).

Telomerase RNP has also been purified from *Euplotes aediculatus*, a hypotrichous ciliate only distantly related to *Tetrahymena* (*12*). The hypotrichs present a special opportunity for telomere studies because their macronuclei contain millions of gene-sized DNA molecules. Each cell has about 8 × 10^7 telomeres (*13*) and about 3 × 10^5 molecules of telomerase (*12*). Measurements of the specific activity of telomerase throughout the purification indicated that the major activity present in macronuclear extracts was purified (*12*). The active telomerase complex had a molecular mass of ~230 kD, corresponding to a 66-kD RNA subunit and two proteins of 123 kD and ~43 kD (*12*). Photocross-linking experiments implicated the larger protein in specific binding of the telomeric DNA substrate (*14*).

Here we characterize the p123 component of *Euplotes* telomerase and show that it contains sequence hallmarks of reverse transcriptases. Furthermore, it is the homolog of a yeast protein, Est2p, shown previously to function in telomere maintenance. Our genetic and biochemical analyses show that the reverse transcriptase motifs of Est2p are essential for telomeric DNA synthesis in vivo and in vitro. We propose that telomerase, frequently called "a specialized reverse transcriptase," is in fact a reverse tran-
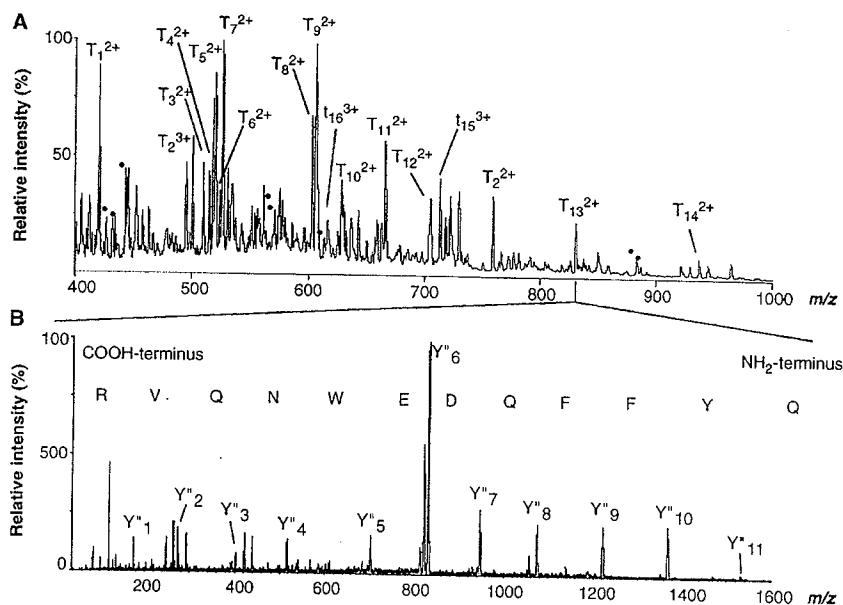
Fig. 1. Sequencing of the p123 subunit of telomerase by nanoelectrospray tandem mass spectrometry. (**A**) Mass spectrum of the unseparated peptide mixture. All peptides that were sequenced completely or partially are marked by the letter T or t, respectively (*15*). The eight peptide ions from which sequence tags were generated are marked by filled circles. Most unlabeled peaks correspond to trypsin autolysis products. (**B**) Tandem mass spectrum of the doubly charged precursor at the mass-to-charge ratio (*m/z*) of 830.4 in (A). Interpretation of the fragment ion mass in (B) and comparison with the esterified form of the peptide allowed the sequence assignment.

J. Lingner and T. R. Cech are at the Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309–0215, USA. T. R. Hughes and V. Lundblad are in the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. A. Shevchenko and M. Mann are at the European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany.

*Present address: Swiss Institute for Experimental Cancer Research, 1066 Epalinges/VD, Switzerland.
†To whom correspondence should be addressed.

scriptase in terms of its catalytic active site.

Determination of Euplotes p123 sequence. The genes encoding the telomerase protein subunits from E. aediculatus were isolated by reverse genetics. Telomerase was purified and polypeptides were separated on SDS-polyacrylamide gels. Amino acid sequencing of the trypsin-digested p123 band was accomplished by nanoelectrospray tandem mass spectrometry (15–17), a miniaturized form of electrospray (18) that allows mass spectrometric interrogation of minute analyte volumes for extended periods of time due to its low flow rate. No chromatography is needed, because the unfractionated peptide mixture obtained after digestion of the protein in a gel slice is separated and sequenced in the spectrometer. For p123, 14 peptides were sequenced de novo (Fig. 1) (15).

Two of the peptide sequences were used to design degenerate polymerase chain reaction (PCR) primers (arrows in Fig. 2) to amplify a portion of the macronuclear gene encoding p123. A genomic library was prepared from macronuclear DNA and screened with this fragment to isolate the full-length gene (19). The p123 gene was found to be encoded by a 3279–base pair (bp) macronuclear chromosome containing an uninterrupted 1031–amino acid open reading frame. In a Southern (DNA) blot experiment the PCR fragment hybrid-

ized to a single macronuclear chromosome of ~3.3 kb (20). The open reading frame predicts a protein of 122,562 daltons, corresponding to the size estimated by SDS-polyacrylamide gel electrophoresis of purified protein [120 kD (12)]. More than 150 amino acids identified in the purified polypeptide by mass spectrometry could be assigned in the open reading frame (Fig. 2). This includes all 14 peptides that were completely sequenced. The tandem mass spectra of 10 additional peptides also

matched the gene sequence through partial sequences or peptide sequence tags (21).

Reverse transcriptase motifs in Euplotes p123 and its yeast homolog Est2p. In a BLAST search of protein databases, Euplotes p123 was found to be most similar to Saccharomyces cerevisiae Est2p ($P = 7 \times 10^{-7}$) and to a group II intron–encoded reverse transcriptase from the cyanobacterium Calothrix ($P = 2 \times 10^{-4}$) (22, 23). Yeast Est2p has a predicted molecular
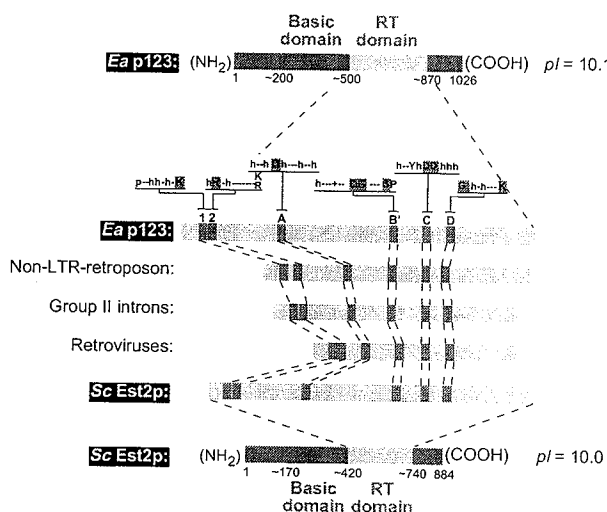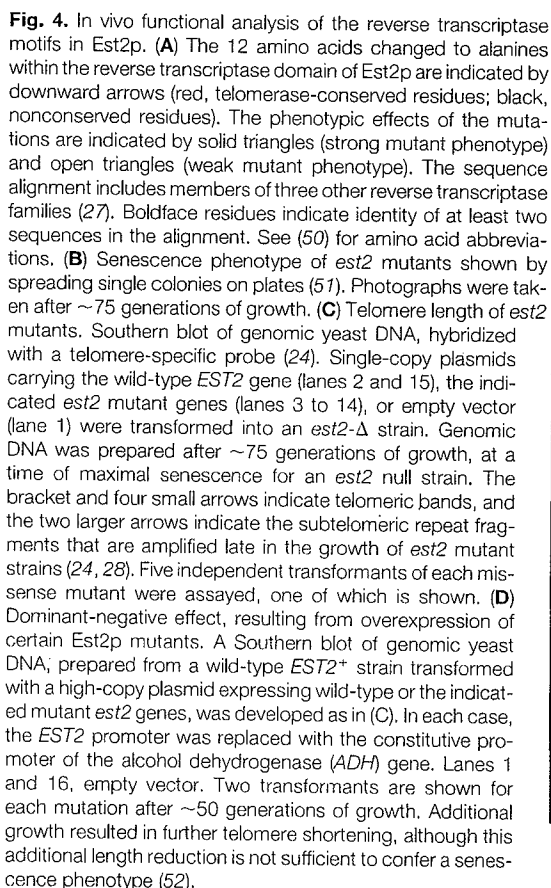

Fig. 3. Block diagrams of p123 and Est2p and comparison of the reverse transcriptase (RT) domains with those of other reverse transcriptases. The spacing of sequence motifs (red) is diagnostic for each reverse transcriptase family (27). In the consensus sequence, abbreviations are as in Fig. 2. The isoelectric point (pI) is the pH at which the protein has no net charge.

Fig. 2. Sequence alignment of Euplotes (Ea) p123 and yeast [S. cerevisiae (Sc)] Est2p (50). Identical amino acids are noted in boldface. The PCR primers used to amplify a portion of the gene are indicated by the arrows. Assigned reverse transcriptase motifs [designated by letters (26) or alternatively by numbers in parentheses (27)] are shown in orange, with the most highly conserved amino acids in red. In the consensus sequences of the motifs, h designates a hydrophobic amino acid, p a polar amino acid, and + a positively charged amino

acid. The underlined sequences in p123 are the 14 peptides completely sequenced by nanoelectrospray tandem mass spectrometry. The dashed lines below the p123 sequence indicate another 10 peptides whose tandem mass spectra matched the sequence. One of the peptides contained an acetylated methionine (solid triangle) at its NH₂-terminus, indicating that it was the NH₂-terminal peptide of the protein. The nucleotide sequence of the Euplotes p123 gene has been deposited in GenBank (accession number U95964).

mass of 103 kD and, like p123, is very basic (Fig. 3). Although the overall sequence identity of *Euplotes* p123 and yeast Est2p is only 20% (Fig. 2), sequence similarity (correspondence of acidic, basic, hydrophobic, and hydrophilic amino acids) can be detected over the entire length of the two proteins.

The *EST2* (ever shorter telomeres) gene was one of four complementation groups identified by screening yeast mutants for reduction in telomere length and senescence (24, 25). Epistasis analysis had indicated that the four *EST* genes function in the same pathway as *TLC1*, the gene encoding the telomerase RNA subunit (6), suggesting that the *EST* genes encode either components of the telomerase or positive regulators of its activity. The homology of yeast Est2p with *Euplotes* p123, the latter isolated because of its physical asso-

Fig. 4. In vivo functional analysis of the reverse transcriptase motifs in Est2p. (**A**) The 12 amino acids changed to alanines within the reverse transcriptase domain of Est2p are indicated by downward arrows (red, telomerase-conserved residues; black, nonconserved residues). The phenotypic effects of the mutations are indicated by solid triangles (strong mutant phenotype) and open triangles (weak mutant phenotype). The sequence alignment includes members of three other reverse transcriptase families (27). Boldface residues indicate identity of at least two sequences in the alignment. See (50) for amino acid abbreviations. (**B**) Senescence phenotype of *est2* mutants shown by spreading single colonies on plates (51). Photographs were taken after ~75 generations of growth. (**C**) Telomere length of *est2* mutants. Southern blot of genomic yeast DNA, hybridized with a telomere-specific probe (24). Single-copy plasmids carrying the wild-type *EST2* gene (lanes 2 and 15), the indicated *est2* mutant genes (lanes 3 to 14), or empty vector (lane 1) were transformed into an *est2-Δ* strain. Genomic DNA was prepared after ~75 generations of growth, at a time of maximal senescence for an *est2* null strain. The bracket and four small arrows indicate telomeric bands, and the two larger arrows indicate the subtelomeric repeat fragments that are amplified late in the growth of *est2* mutant strains (24, 28). Five independent transformants of each missense mutant were assayed, one of which is shown. (**D**) Dominant-negative effect, resulting from overexpression of certain Est2p mutants. A Southern blot of genomic yeast DNA, prepared from a wild-type *EST2⁺* strain transformed with a high-copy plasmid expressing wild-type or the indicated mutant *est2* genes, was developed as in (C). In each case, the *EST2* promoter was replaced with the constitutive promoter of the alcohol dehydrogenase (*ADH*) gene. Lanes 1 and 16, empty vector. Two transformants are shown for each mutation after ~50 generations of growth. Additional growth resulted in further telomere shortening, although this additional length reduction is not sufficient to confer a senescence phenotype (52).

ciation with telomerase RNA and its co-purification with telomerase activity, supported the proposal that both proteins are intrinsic subunits of their respective telomerases.

Euplotes p123 contains reverse transcriptase motifs, and the alignment reveals the presence of these motifs in a similar region of Est2p (Fig. 3). The primary sequences of reverse transcriptases are highly divergent: Only a few amino acids are absolutely conserved within separate short motifs (26, 27), but these motifs are believed to form a common tertiary fold. Both p123 and Est2p contain these key conserved amino acids, most notably the three invariant aspartates in motifs A and C, which are thought to be directly involved in catalysis (Fig. 2). Conserved motifs are spaced differently in the two major branches of reverse transcriptases, those encoded by retroviruses and long terminal repeat (LTR) retroposons and those encoded by non-LTR retroposons and group II introns (27). The spacing of sequence motifs in p123 and Est2p resembles that in the latter branch. However, the interval between motifs A and B' in p123 and Est2p is unusually large (Fig. 3), suggesting that these two polypeptides may be members of a previously unknown subcategory.

Requirement of the reverse transcriptase motifs for Est2p function in vivo. The presence of reverse transcriptase motifs in both p123 and Est2p suggests that this region

may define the catalytic active site of telomerase. To test the importance of these motifs for Est2p function, we used site-directed mutagenesis to change conserved and non-conserved aspartic acid (D) and glutamine (Q) residues in and around motifs A, B', and C to alanine (A) (Fig. 4A). Each mutant, present on a single-copy ARS CEN plasmid, was tested for in vivo function in a complementation assay. Plasmids were transformed into the *est2-Δ* strain (Δ designates deletion), in parallel with either the empty vector or an *EST2⁺* plasmid. Transformants were assessed for the senescence phenotype (Fig. 4B) and for chromosome telomere length (Fig. 4C).

Consistent with the prediction that the reverse transcriptase motifs are required for Est2p function, mutation of any of the three conserved aspartates in motifs A and C prevented normal telomerase activity. Transformants expressing these mutant proteins became senescent and had shortened telomeric tracts, phenotypes indistinguishable from those of the null mutant (Fig. 4, B and C). Furthermore, a bypass pathway for telomere maintenance (28) was evident in these three mutant strains. Activation of this alternative pathway occurs as the result of a global amplification and rearrangement of both telomeric G-rich repeats and subtelomeric regions, and has only been observed in *est* and *tlc1* mutant strains with a severe telomere shortening phenotype (24, 29). A feature of this pathway is the amplification
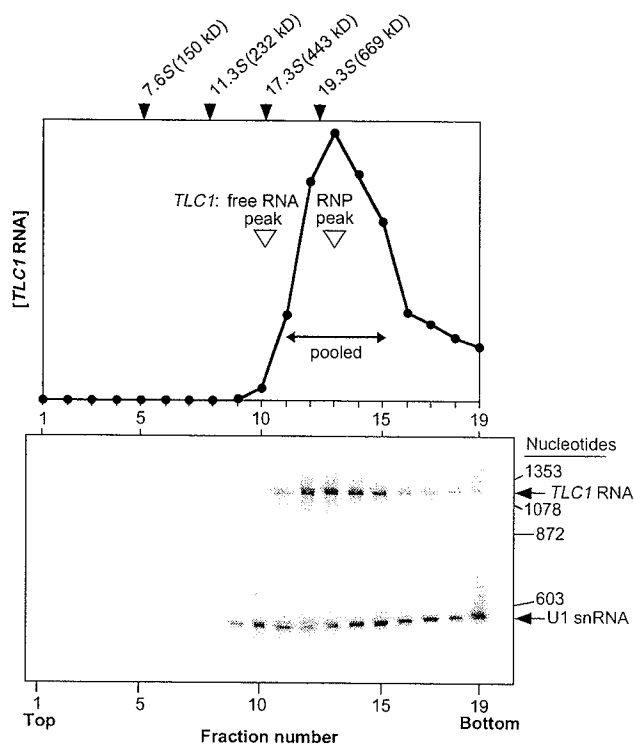
of two subtelomeric bands (Fig. 4C); these diagnostic restriction fragments were substantially amplified only in the *est2* null mutant and the three proposed active site mutants.

Mutations of amino acids other than the three most conserved aspartates had less severe or no phenotypic effects. The residue Asp⁵³⁶ of motif A is conserved between Est2p and p123, and the D536A mutation (Asp mutated to Ala at position 536) caused substantial telomere shortening and a modest senescence phenotype. Of the conserved residues tested, Gln⁶³² of motif B' was the only one that was functionally insensitive to replacement with alanine. However, this glutamine is not strictly conserved in reverse transcriptases (27), and when it is changed to alanine in human immunodeficiency virus–1 (HIV-1) reverse transcriptase, polymerase activity in vitro is reduced but not completely eliminated (30). In contrast to the phenotypes seen upon mutation of the semiconserved amino acids, mutation of six of the seven nonconserved amino acids tested showed little or no alteration of Est2p function.

Two observations indicate that stable Est2 protein was produced in the five *est* mutants with a diminished capacity to complement the *est2-Δ* strain. First, Myc₃-epitope-tagged versions of each mutant protein were visualized immunologically after immunoprecipitation (31). Second, overexpression of each of the five mutant alleles in a wild-type yeast strain with a functional chromosomal *EST2⁺* gene resulted in telomere shortening (Fig. 4D), whereas overexpression of the wild-type *EST2* gene had little effect. The dominant-negative phenotype shows that each mutant protein is being made and suggests that excess mutant Est2p can titrate components away from the wild-type telomerase complex.

Requirement of Est2p for telomerase activity in vitro. If Est2p is the catalytic protein subunit of telomerase, then telomerase activity should be abolished in *est2* mutant extracts. An in vitro assay was developed with extracts fractionated by glycerol gradient centrifugation (32). Telomerase-containing fractions were identified by detection of the RNA subunit on Northern blots (Fig. 5). Yeast telomerase sedimented as a 19S to 20S particle, substantially faster than the sedimentation of the deproteinized telomerase RNA (~17S). Telomerase-containing glycerol gradient fractions were pooled, concentrated, and tested for the ability to elongate a single-stranded telomeric oligonucleotide (Fig. 6). An activity was detected in wild-type extracts that had the characteristics of telomerase. It was dependent on the presence of oligonucleotide substrate and fractionated extract (Fig. 6A,

Fig. 5. Sedimentation of telomerase. Yeast extract was fractionated on a glycerol gradient (32), and telomerase RNA was detected by Northern blotting (bottom) and its concentration quantified on a Phosphorimager (top). Detection of U1 snRNP served as an internal control. Fractions pooled for activity assays are indicated. Telomerase RNP sedimented as a 19S to 20S particle, whereas deproteinized telomerase RNA sedimented at ~17S. The sedimentation value was determined relative to marker proteins that were run in parallel gradients and that had sedimentation coefficients of 7.6S (alcohol dehydrogenase), 11.3S (catalase), 17.3S (apoferritin), and 19.3S (thyroglobulin).

lanes 1 to 3). Addition of T and G residues occurred in an ordered manner consistent with the expected alignment of substrate and RNA template (Fig. 6A, lanes 5 and 6). The activity was sensitive to low concentrations of ribonuclease (RNase) A and was not stimulated by adenosine triphosphate (ATP) (Fig. 6B). These characteristics, in addition to the observed single round of extension of primer (Fig. 6A), are similar to those of the telomerase activity described by Blackburn and co-workers (33). A different activity described as telomerase by Lue and Wang (34) gives rise to long products and is stimulated by ATP. This latter activity was not detectable in our telomerase-containing glycerol gradient fractions.

A telomerase RNA template mutation that alters the specificity of nucleotide incorporation to produce a Hae III restriction site (6) provides an additional test for the authenticity of the in vitro telomerase assay. An extract of this TLC1-1(HaeIII) mutant, fractionated on a glycerol gradient, gave the predicted extension of the telomeric oligonucleotide only in the presence of deoxycytidine triphosphate (dCTP) (Fig. 6C, lanes 6 to 8), a nucleotide that has no effect on extension by a wild-type extract (Fig. 6C, lanes 2 and 3). This nucleotide specificity change supports the dependence of the assay on the TLC1 RNA. Because the TLC1-1(HaeIII) strain also undergoes senescence (29), this result also provides confidence that telomerase activity can still be detected in senescing cells, as long as they are not subcultured too extensively.

We then assayed fractionated extracts from est2-Δ and tlc1-Δ strains for telomerase activity (Fig. 6D). As expected, no activity was detectable in tlc1-Δ yeast, which has the gene for telomerase RNA deleted. In the est2-Δ strain, telomerase RNA was still assembled into an RNP, as assessed by glycerol gradient centrifugation and Northern blotting (32), but telomerase activity was completely absent. This indicates that Est2p is essential for telomerase activity. As described above, the absence of activity is not simply a secondary consequence of senescence. We also measured telomerase activity in extracts from est2-Δ and strains expressing two of the proposed active site mutants in the presence of the chain-terminating analog ddGTP (Fig. 6E). According to the proposed primer-template alignment, extension should terminate after addition of two nucleotides. A practical advantage is the higher signal-to-noise ratio obtained when all products are concentrated in one or two bands. Again, activity was dependent on functional TLC1 and EST2 genes.

Telomerase structure. The presence of a reverse transcriptase domain in the catalytic subunit of telomerase provides a framework

for exploring the structure and mechanism of this enzyme. Reverse transcriptases have been studied in great detail, and the three-dimensional structure of HIV-1 reverse transcriptase has been solved (35). The structure can be compared with a right hand with fingers, palm, and thumb, with the active site residing in the palm (36). A model for telomerase structure based on that of HIV-1 reverse transcriptase (HIV-1 RT) is shown in Fig. 7 with the telomerase RNA and a telomeric DNA substrate superimposed.

The catalytic subunit of telomerase has several features that distinguish it from other reverse transcriptases. Telomerase uses only a small portion of its RNA subunit as a template. The borders of this template

must somehow be recognized. Furthermore, during processive synthesis of telomeric repeats the substrate translocates from one end of the template to the other by an as yet unknown mechanism. The large gap between motifs A and B' of telomerase p123 and Est2p indicates an unusual finger domain structure. In HIV-1 RT this domain may be involved in template strand binding (35, 36); whether and how it contributes to the unusual reaction mechanism of the telomerase RNP remain to be investigated. Finally, the telomerase protein is stably associated with its RNA subunit, as shown by our isolation of the Euplotes p123-RNA complex and by coimmunoprecipitation of the yeast RNA subunit with Est2p (31).
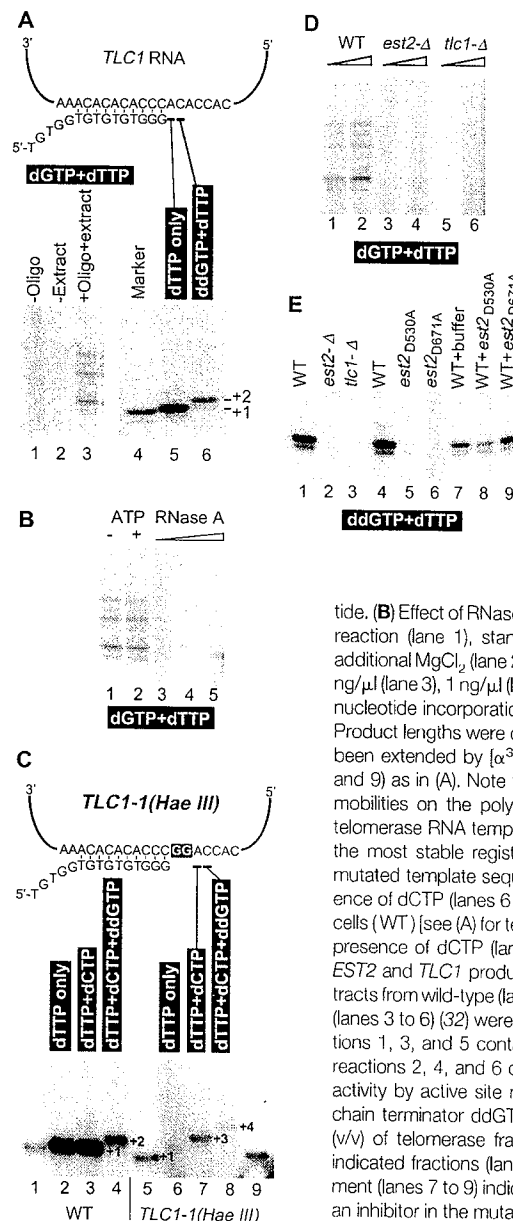

Fig. 6. In vitro functional analysis of reverse transcriptase motifs in Est2p. Telomerase was partially purified by glycerol gradient centrifugation and assayed for the ability to extend a telomeric DNA substrate (32). In the assay $[\alpha^{32}P]$dTTP was included to visualize products elongated by 1, 2, 3, or 4 nucleotides (+1, +2, +3, or +4). (**A**) The telomerase RNA template region maximally base-paired to the DNA substrate is indicated schematically. Product lengths were determined relative to the same DNA substrate extended by one nucleotide at its 3' end by reaction with $[\alpha^{33}P]$ddTTP and terminal deoxynucleotidyl transferase (lane 4). Up to seven nucleotides were added in the presence of dGTP and dTTP (lane 3), one nucleotide in the presence of only dTTP (lane 5), and two nucleotides in the presence of dTTP and the chain-terminating analog ddGTP (lane 6). Oligo, oligonucleotide. (**B**) Effect of RNase A and ATP on telomerase activity. Standard reaction (lane 1), standard reaction plus 1 mM ATP and 1 mM additional MgCl$_2$ (lane 2), and standard reaction plus RNase A at 0.1 ng/μl (lane 3), 1 ng/μl (lane 4), and 10 ng/μl (lane 5). (**C**) Specificity of nucleotide incorporation dictated by the RNA template sequence. Product lengths were determined relative to DNA markers that had been extended by $[\alpha^{32}P]$ddTTP (lane 1) or $[\alpha^{32}P]$ddCTP (lanes 5 and 9) as in (A). Note that these two markers had slightly different mobilities on the polyacrylamide gel. The mutant TLC1-1(HaeIII) telomerase RNA template is indicated with the substrate bound in the most stable register. Consistent with this alignment and the mutated template sequence, efficient extension required the presence of dCTP (lanes 6 and 7). Telomerase in extract from TLC1-wt cells (WT) [see (A) for template sequence] was not influenced by the presence of dCTP (lanes 2 and 3). (**D**) Requirement of functional EST2 and TLC1 products for telomerase activity. Fractionated extracts from wild-type (lanes 1 and 2) and the indicated mutant strains (lanes 3 to 6) (32) were tested at two extract concentrations. Reactions 1, 3, and 5 contained 10% (v/v) of telomerase fraction, and reactions 2, 4, and 6 contained 20%. (**E**) Alleviation of telomerase activity by active site mutations in Est2p. All assays included the chain terminator ddGTP (100 μM). The reactions contained 10% (v/v) of telomerase fraction (lanes 1 to 6) or 5% of each of the indicated fractions (lanes 7 to 9). The results of the mixing experiment (lanes 7 to 9) indicate that the absence of activity is not due to an inhibitor in the mutant extracts.

This last feature distinguishes telomerase from the retroviral and LTR retroposon reverse transcriptases, but is similar to some mitochondrial and group II intron–encoded reverse transcriptases that also form complexes with their RNA templates (37).

Reverse transcriptase essential for chromosome replication in diverse eukaryotes. Reverse transcriptases have not previously been considered essential for normal cell physiology. Initially discovered as retroviral enzymes that catalyze the defining RNA-to-DNA step of retroviral replication (38), they were later found to mediate the transposition of DNA elements within eukaryotic genomes through an RNA intermediate (39). Reverse transcriptases are also present in some prokaryotes (40) and in *Neurospora* mitochondria (41), where they replicate genetic elements that are nonessential to their "host." Our discovery that a structurally related enzyme is essential for chromosome replication and cell division provides another example of the opportunism of nature: once a useful protein motif is stumbled upon, natural selection promotes its exploitation in diverse ways.

The evolutionary relationship between telomerase and the other reverse transcriptases is intriguing. It is well established that retroviruses acquired oncogenes such as v-*src*, v-*abl*, v-*ras*, and v-*fos* from cellular genomes. According to Temin's protovirus hypothesis, retroviruses also acquired their reverse transcriptase gene from normal cells, where the enzyme presumably contributed to some normal cellular process (42). Could this cellular source have been the telomerase p123/EST2 gene, which mutated so that the protein product used an

exogenous rather than an intrinsic RNA template? Alternatively, telomerase and the reverse transcriptases encoded by retrotransposons and retroviruses may all be descendants of an ancestral protein that emerged from an "RNA world" (43).

Telomere replication in the fruit fly *Drosophila* has been mysterious because this organism does not have short repeated telomeric sequences and presumably no telomerase. Rather, the non-LTR retroposons *HeT-A* and *TART* cap the chromosome ends (44). The *TART* reverse transcriptase is closely related to p123 and Est2p, which suggests that the *Drosophila* telomere replication machinery may in fact not be so different from that of other eukaryotes (45).

We have no satisfactory explanation for the lack of correspondence between the *Euplotes* and yeast p123/Est2p proteins and the *Tetrahymena* p80 or p95 protein (9). The small protein subunit of *Euplotes* telomerase (p43) also shows no similarity to the *Tetrahymena* proteins (46), and the complete yeast genome sequence does not reveal obvious p80 and p95 homologs. There are three possible explanations: (i) *Tetrahymena* may have a different telomerase in which p80 and p95 provide the active site (that is, telomerase was invented more than once in evolution). (ii) *Tetrahymena* may have two telomerases, one containing p80 and p95 and one (unisolated) containing a p123/Est2p homolog (for example, one telomerase for de novo telomere formation during macronuclear development and one for telomere replication). (iii) The *Tetrahymena* p80-p95-RNA complex may not be an active enzyme but may require a p123/Est2p subunit that was underrepresented upon purification of the particle.
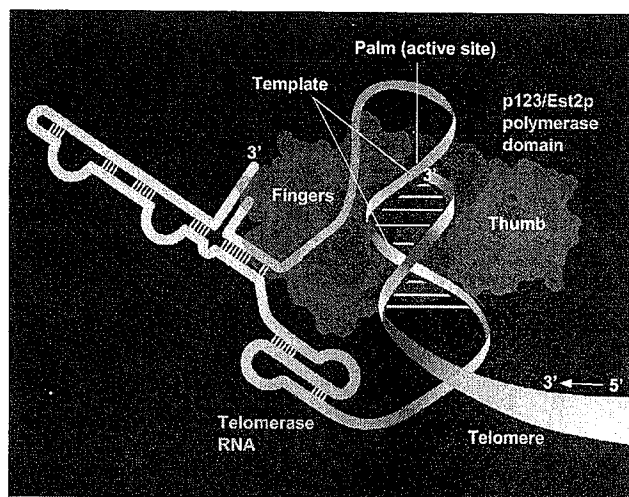
Mass spectrometric methods have recently become very successful for the identification of proteins whose genes are already partially or completely contained in sequence databases (47). The sequencing of more than 150 amino acids of the p123 telomerase subunit at protein amounts too low for chemical methods shows that mass spectrometry is now also valuable for sequencing previously unidentified proteins.

Telomerase activation accompanies the immortalization of cultured mammalian cells and is also a common property of human tumor cells (48). Thus, telomerase is considered to be a potential target for the development of tumor-specific drugs. Certain reverse transcriptase inhibitors developed as anti-HIV drugs have already been tested against telomerase with some success (49). The finding that the telomerase active site is related to that of known reverse transcriptases is expected to stimulate such efforts.

REFERENCES AND NOTES

1. J. D. Watson, *Nature* 239, 197 (1972); J. Olovnikov, *J. Theor. Biol.* 41, 181 (1973).
2. J. Lingner, J. P. Cooper, T. R. Cech, *Science* 269, 1533 (1995); for a similar model regarding the macronuclear DNA of hypotrichous ciliates, see A. M. Zahler and D. M. Prescott, *Nucleic Acids Res.* 16, 6953 (1988).
3. C. W. Greider and E. H. Blackburn, *Cell* 51, 887 (1987).
4. _____, *Nature* 337, 331 (1989); D. Shippen-Lenz and E. H. Blackburn, *Science* 247, 546 (1990); D. P. Romero and E. H. Blackburn, *Cell* 67, 343 (1991).
5. J. Lingner, L. L. Hendrick, T. R. Cech, *Genes Dev.* 8, 1984 (1994).
6. M. S. Singer and D. E. Gottschling, *Science* 266, 404 (1994).
7. M. J. McEachern and E. H. Blackburn, *Nature* 376, 403 (1995).
8. J. Feng *et al.*, *Science* 269, 1236 (1995); M. A. Blasco, W. Funk, B. Villeponteau, C. W. Greider, *ibid.*, p. 1267.
9. K. Collins, R. Kobayashi, C. W. Greider, *Cell* 81, 677 (1995).
10. L. Harrington *et al.*, *Science* 275, 973 (1997); J. Nakayama, M. Saito, H. Nakamura, A. Matsuura, F. Ishikawa, *Cell* 88, 875 (1997).
11. J. Lingner and T. Nakamura, unpublished data. The BLAST search of protein databases was performed at the National Center for Biotechnology Information with GCG (Program Manual for the Wisconsin Package, Version 8.1, Genetics Computer Group, Madison, WI).
12. J. Lingner and T. R. Cech, *Proc. Natl. Acad. Sci. U.S.A.* 93, 10712 (1996).
13. D. M. Prescott, *Microbiol. Rev.* 58, 233 (1994).
14. P. Hammond, T. Lively, T. R. Cech, *Mol. Cell. Biol.* 17, 296 (1997).
15. Amino acid analysis of 10% of the protein indicated a low picomole amount, insufficient for internal sequencing by Edman degradation. Protein bands were excised from the Coomassie-stained gel and digested in the gel (16). Reconstituted peptide mixture was desalted on a 50-nl microcartridge and eluted directly into the nanoelectrospray needle. Peptides were sequenced on a tandem mass spectrometer API III (PE-Sciex, Ontario, Canada) equipped with a nanoelectrospray ion source (17). Mass spectra (Q1 scans) were recorded with 0.1-dalton stepwidth and unit mass resolution. For the tandem mass spectrometric investigation, Q1 was set to transmit a 2-dalton mass window. The tan-

Fig. 7. Model of telomerase as an RNA–reverse transcriptase complex. The p123/Est2p subunit (green) is based on the right hand model of HIV-1 RT (35); thumb and fingers extend toward the reader. Motifs A, B', C, and D are in the palm, and the active site aspartates are near the 3' end of the telomeric DNA substrate (red). The RNA subunit (purple) has its template region in the palm; the location of the remainder of the RNA is unknown and is shown schematically in its secondary structure representation (5). Additional protein subunits may be associated (not shown). The telomeric DNA substrate is shown base-paired but not intertwined with the RNA subunit. The extent of base pairing and the sites of interaction of the nucleic acids with the protein are not known.

dem mass spectra were acquired with 0.2-dalton stepwidth. The mass resolution was sufficient to assign unambiguously an amino acid series to the identifiable COOH-terminal ion series. The procedure was repeated with the esterified peptide mixture as described (16). The following peptide sequences (50) were assigned: T1, YLLFQR; T2, DYNEEDFQVL(VK,AR); T3, (L,N)WDVLMK; T4, SFLMNNLTHYFR; T5, TLYSWLQK; T6, ETLAE-VQEK; T7, AMLGNELFR; T8, LQTSFPLSPSK; T9, (A,Q)TLFTNLFR; T10, (L,T)ALMPNLNLR; T11, LFAQTNLVATPR; T12,LESWMQVETSAK; T13, QYFFQDEWNQVR; and T14, (E,D)SMNPENPNVN-LLMR. The partial sequence EVDVD was obtained in the NH$_2$-terminal peptide t15 (MEVDVD-NQADNHGIHSALK) and the sequence LFFAT in the peptide t16 (LFFATMDIEK). There are two discrepancies between the sequence determined by mass spectrometry and that deduced from the DNA sequence. The glutamine in peptide T2 was a lysine, and the order of the first two amino acids in T12 was reversed. In the above sequences, L signifies isoleucine or leucine, two isobaric amino acids that cannot be distinguished by our methods, and the parentheses indicate that the order of the first two amino acids could not be distinguished.

16. M. Wilm et al., Nature 379, 466 (1996).
17. M. S. Wilm and M. Mann, Int. J. Mass Spectrom. Ion Proc. 136, 167 (1994); Anal. Chem. 68, 1 (1996).
18. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Science 246, 64 (1989).
19. DNA was isolated from E. aediculatus, treated with mung bean nuclease, and subcloned into the Sma I site of pCR-Script SK (+) (Stratagene). The macronuclear gene encoding p123 was identified by colony hybridization.
20. J. Lingner and T. R. Cech, unpublished data.
21. M. Mann and M. S. Wilm, Anal. Chem. 66, 4390 (1994).
22. P is the probability that the specified segment pair score would be observed in a search of the same size with a random sequence.
23. J.-L. Ferat and F. Michel, Nature 364, 358 (1993).
24. T. S. Lendvay, D. K. Morris, J. Sah, B. Balasubramanian, V. Lundblad, Genetics 144, 1399 (1996).
25. EST1, which has been described [V. Lundblad and J. W. Szostak, Cell 57, 633 (1989)], is not essential for core telomerase activity (33). EST4 is identical to CDC13 [B. Garvik, M. Carson, L. Hartwell, Mol. Cell. Biol. 15, 6128 (1995)]. Biochemical analysis has shown that Est1p and Cdc13p are single-stranded telomeric DNA binding proteins that are proposed to facilitate access of telomerase to chromosome ends [C. I. Nugent, T. R. Hughes, N. F. Lue, V. Lundblad, Science 274, 249 (1996); J. J. Lin and V. A. Zakian, Proc. Natl. Acad. Sci. U.S.A. 93, 13760 (1996); V. Virta-Pearlman, D. K. Morris, V. Lundblad, Genes Dev. 10, 3094 (1996)].
26. O. Poch, I. Sauvaget, M. Delarue, N. Tordo, EMBO J. 8, 3867 (1989).
27. Y. Xiong and T. H. Eickbush, ibid. 9, 3353 (1990).
28. V. Lundblad and E. H. Blackburn, Cell 73, 347 (1993).
29. C. Nugent and V. Lundblad, unpublished data.
30. S. G. Sarafianos, V. N. Pandey, N. Kaushik, M. J. Modak, Biochemistry 34, 7207 (1995).
31. Est2p was overexpressed under control of the ADH promoter as a Myc$_3$-tagged fully functional derivative. Extracts were immunoprecipitated with an antibody to c-Myc. Wild-type and mutant Est2 proteins were precipitated with the antibody and detected on an immunoblot. Telomerase RNA was coimmunoprecipitated with Est2p only in strains harboring the Myc$_3$-tagged gene and was detected on a Northern (RNA) blot. U1 snRNA was not substantially precipitated but was detected in the crude extract.
32. Yeast was grown in yeast extract, peptone, and dextrose (YPD) medium to an optical density (600 nm) of 1.8, and extracts were prepared as described [A. Ansari and B. Schwer, EMBO J. 14, 4001 (1995)] with the following modifications: Cell pellets were resuspended in two volumes of extract buffer [20 mM tris-acetate (pH 7.5), 300 mM potassium glutamate, 1.1 mM MgCl$_2$, 0.1 mM EDTA, 5% glycerol, 1 mM dithiothreitol (DTT), and 0.5 mM phenylmethylsulfonyl fluoride]. Extracts were concentrated threefold on centricon 30 (Amicon) and fractionated by glycerol gradient centrifugation in extract buffer containing 15 to 40% glycerol (12). Centrifugation was performed in a SW55Ti rotor (Beckman) at 304,000g_{max} for 12 hours at 4°C. Nineteen fractions were collected from the top of the gradient and assayed for the telomerase RNA subunit TLC1 by Northern blotting. In est2-Δ and est2 mutant strains, telomerase did not show a major shift in its sedimentation coefficient relative to the wild type, but the large size of the RNA and the limited resolution of the glycerol gradients (Fig. 5) would have prevented a small shift from being observed. Telomerase-containing fractions were pooled and concentrated fivefold on centricon 30. Telomerase assays (10 μl) contained 20 mM tris-acetate (pH 7.5), 1 mM MgCl$_2$, 30 mM potassium glutamate, 1 mM DTT, 1 μM oligonucleotide substrate (5'-TGTGGTGTGTGTGGG-3'), 100 μM deoxy- or dideoxyguanosine triphosphate (dGTP or ddGTP), 15% (v/v) [α-^{32}P]dTTP (800 Ci/mmol; 10 μCi/μl), and 10% (v/v) telomerase fraction. After 1 hour at 30°C, the reactions were stopped, digested with proteinase K, ethanol precipitated, and separated as in (5). Fractions were derived from the following strains: Fig. 6, A and B: TVL268 [Mata ura3-52 ade2-101 lys2-801 leu2-Δ1 his3-Δ200/CF (TRP1 SUP11)]; Fig. 6C, lanes 2 to 4: TVL268; Fig. 6C, lanes 6 to 8: TVL268 TLC1-1(HaeIII); Fig. 6D, lanes 1 and 2, and Fig. 6E, lane 1: AVL78 (Mata leu2 trp1 ura3-52 prb prc pep4-1); Fig. 6D, lanes 3 and 4, and Fig. 6E, lane 2: AVL78 est2-Δ1::URA3; Fig. 6D, lanes 5 and 6, and Fig. 6E, lane 3: AVL78 tlc1-Δ::LEU2; Fig. 6E, lanes 4, 7, 8, and 9: TVL268; Fig. 6E, lanes 5 and 8: TVL268 est2-D530A; and Fig. 6E, lanes 6 and 9: TVL268 est2-D671A. The est2-Δ1 and tlc1-Δ deletion strains were constructed by one-step gene disruption, whereas the est2-D530A, est2-D671A, and TLC1-1(HaeIII) mutant alleles were constructed by replacement of the EST2-wt and TLC1-wt wild-type alleles, respectively. In each case, extracts were prepared for analysis after as few generations of growth as possible.
33. M. Cohn and E. H. Blackburn, Science 269, 396 (1995); J. Prescott and E. H. Blackburn, Genes. Dev. 11, 528 (1997).
34. N. F. Lue and J. C. Wang, J. Biol. Chem. 270, 21453 (1995).
35. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, Science 256, 1783 (1992); E. Arnold et al., Nature 357, 85 (1992); A. Jacobo-Molina et al., Proc. Natl. Acad. Sci. U.S.A. 90, 6320 (1993).
36. D. L. Ollis, P. Brick, T. A. Steitz, Nature 313, 765 (1985). For comparison of different polymerase structures, see C. M. Joyce and T. A. Steitz, Annu. Rev. Biochem. 63, 777 (1994), and R. Sousa, Trends Biochem. Sci. 21, 186 (1996).
37. J. C. Kennell, H. Wang, A. M. Lambowitz, Mol. Cell. Biol. 14, 3094 (1994); S. Zimmerly et al., Cell 83, 529 (1995).
38. D. Baltimore, Nature 226, 1209 (1970); H. M. Temin and S. Mizutani, ibid., p. 1211.
39. J. D. Boeke and V. G. Corces, Annu. Rev. Micro. 43, 403 (1989).
40. M. Inouye and S. Inouye, ibid. 45, 163 (1991).
41. H. Wang and A. M. Lambowitz, Cell 75, 1071 (1993).
42. H. M. Temin, Annu. Rev. Genet. 8, 155 (1974).
43. J. E. Darnell and W. F. Doolittle, Proc. Natl. Acad. Sci. U.S.A. 83, 1271 (1986). See also N. Maizels and A. Weiner, in The RNA World, R. F. Gesteland and J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993), pp. 577–602.
44. H. Biessmann et al., Cell 61, 663 (1990); R. W. Levis, R. Ganesan, K. Houtchens, L. A. Tolar, F. Sheen, ibid. 75, 1 (1993).
45. M. L. Pardue, O. N. Danilevskaya, K. Lowenhaupt, F. Slot, K. L. Traverse, Trends Genet. 12, 48 (1996).
46. J. Lingner, A. Shevchenko, M. Mann, T. R. Cech, unpublished data.
47. See, for example, A. Shevchenko et al., Proc. Natl. Acad. Sci. U.S.A. 93, 14440 (1996). Reviewed in S. D. Patterson and R. Aebersold, Electrophoresis 16, 791 (1995).
48. C. M. Counter et al., EMBO J. 11, 1921 (1992); T. de Lange, Proc. Natl. Acad. Sci. U.S.A. 91, 2882 (1994).
49. C. Strahl and E. H. Blackburn, Nucleic Acids Res. 22, 893 (1994); Mol. Cell. Biol. 16, 53 (1996).
50. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
51. Twenty single colonies from five independent transformations of an est2-Δ strain were assayed for each mutation. The results, including the slight reduction in colony size observed for est2$_{D536A}$ and est2$_{D592A}$, were highly reproducible. In addition, the senescence phenotype of these two mutants (as well as that of the more severely affected est2$_{D530A}$, est2$_{D670A}$, and est2$_{D671A}$ mutants) was enhanced in an est2 rad52 strain background (29). The absence of RAD52 gene function has been shown to eliminate a backup pathway for telomere maintenance, thereby enhancing the phenotype of est mutant strains (24, 28).
52. S. Evans and V. Lundblad, unpublished data.
53. We thank C. Grosshans, K. Goodrich, and N. Walcott for technical assistance; T. Nakamura, C. Chapon, A. Zaug, and J. Hansen for many helpful discussions; and D. Underwood and E. Jabri for help with Fig. 7. Supported in part by the Swiss National Science Foundation (J.L.), NIH grant AG11728 and Geron Corporation (V.L.), and a grant from the German Technology Ministry (Bundesministerium für Biologische Forschung) (M.M.).

4 March 1997; accepted 2 April 1997

Functional analysis of HIV-1 reverse transcriptase motif C: site-directed mutagenesis and metal cation interaction.

Valverde-Garduno V, Gariglio P, Gutierrez L.

Departamento de Virus y Cancer, Instituto Nacional de Salud Publica, Cuernavaca, Morelos, Mexico. veronica@ibt.unam.mx

Motif C, present in all polymerases, has been proposed to be part of the catalytic and metal binding site of the enzyme, suggesting that polymerases have a common origin. Previously, we have shown that the metal ion manganese induces alterations in nucleotide substrate specificity in some polymerases. However, it is not known if the active site responsible for incorporation of nonspecific substrates is the same as that which incorporates specific ones. Here we show that manganese enables HIV-1 reverse transcriptase (RT) to incorporate rNTP's using RNA as a template, thus behaving as an RNA replicase. Also, we show that the mutation D186H in motif C strongly affects the natural DNA polymerase activity and that the RNA replicase activity becomes undetectable, suggesting that both activities depend on the same active site. This mutation changes the metal ion preference, with mutant RT presenting only 0.5% of the wild-type DNA polymerase activity in the presence of magnesium but 1.6% of the same activity in the presence of manganese. This variation in cation preference suggests that residue D186 is part of the metal binding site. Since residue D186 of motif C is essential for both activities and appears to be involved in the binding of an important cation needed for the specific activity, our results support the idea of a common origin for all polymerases, from an ancestral unspecified polymerase containing at least motif C.

Rapid Identification of All Known Retroviral Reverse Transcriptase Sequences with a Novel Versatile Detection Assay

WOLFGANG SEIFARTH,[1] ULF KRAUSE,[1] CHRISTINE HOHENADL,[1,2] CORINNA BAUST,[1]
RÜDIGER HEHLMANN,[1] and CHRISTINE LEIB-MÖSCH[1,2]

ABSTRACT

We have developed a highly sensitive, universal assay that allows detection as well as identification of all known retroviral reverse transcriptase (RT)-related nucleic acids in a biological sample by a single two-step experiment. The assay combines polymerase chain reaction (PCR) and reverse dot-blot hybridization (RDBH), using an array of immobilized synthetic retrovirus-specific oligonucleotides and two sets of mixed oligo primers (MOPs). These primers were derived from highly conserved motifs found in all known reverse transcriptase genes. The PCR/RDBH assay was used for qualitative analyses of human endogenous retrovirus (HERV) transcription in peripheral blood mononuclear cells (PBMCs) and in particles released by the human mammary carcinoma-derived cell line T47D. Sensitivity was further demonstrated by detection of down to 10 copies of pig endogenous retrovirus (PERV) DNA in human cDNA samples. Therefore, this assay is particularly useful for the identification of retroviral sequences in xenografts as well as in recipients of xenografted tissues and organs. Moreover, it is a valuable tool to detect retroviral transcripts and particles in cell cultures used for production of therapeutic polypeptides. The assay is further suitable for monitoring vector preparation used in human gene therapy to exclude transfer of copackaged endogenous retroviruses into target cells.

INTRODUCTION

THE GENOMES OF ALL VERTEBRATES contain a wide spectrum of endogenous retroviruses (ERVs) and reverse transcriptase (RT)-related sequences. For example, human ERVs (HERVs) are estimated to comprise at least 1–2% of the human genome.[1,2] Although most of these sequences are assumed to be defective, some retain certain biological activities and thus represent a reservoir of retroviral genes with pathogenic potential. Characterization of particles released by the human breast cancer-derived cell line T47D revealed that complementation between several expressed HERVs can lead to pseudotype particles packaging retroviral RNA of different origin.[3,4] Thus, activation and expression of (H)ERV may result in undesired mobilization of genetic material of retroviral origin and may interfere with the safe production of therapeutic polypeptides, with safe human gene therapy and xenotransplantation.[5]

Cross-packaging of ERVs to a high level is observed in murine packaging cells commonly used for retroviral vector preparation.[6] Cross-packaged ERV transcripts may be transmitted to recipient cells leading to unwanted integration events, or may recombine with the vector forming new infectious retroviruses. This is of high concern for the safety of retrovirus-mediated human gene therapy. The risk of acquiring animal ERVs through xenotransplantation also requires attention since xenografts from baboons and pigs are currently discussed for human use. In vitro experiments indicate that xenotropic ERVs such as murine or cat retroviruses can propagate considerably in human cells.[7] Baboon endogenous retrovirus (BaEV) read-

ily infects human cells in culture. The same holds true for porcine ERV and several human cell types.[5,8-10] Although these viruses do not show a pathogenic effect in their natural hosts, the situation may change when they are transferred to immunosuppressed humans, in whom the virus might replicate to high titers.[11,12]

With respect to prospective practical applications we have established a universal detection assay that allows rapid testing of biological samples for undesired mobilization of retroviral sequences. With this method all known reverse transcriptase-related sequences of human and animal origin can be simultaneously identified in a single two-step experiment.

MATERIALS AND METHODS

RNA preparation

Total RNA was extracted from peripheral blood mononuclear cells (PBMCs) of healthy blood donors according to a guanidinium isothiocyanate–cesium chloride (GIT/CsCl) ultracentrifugation protocol[13] and dissolved in diethylpyrocarbonate (DEPC)-treated distilled water. Consecutively, mRNA was purified with Dynabeads paramagnetic particles as described by the manufacturer (Dynal, Hamburg, Germany). Nucleic acid concentrations were calculated by spectrophotometry at 260 nm. To check for genomic DNA contaminations 50 ng of each mRNA preparation was tested in a polymerase chain reaction (PCR) with mixed oligonucleotide (oligo) primers (MOPs) omitting the reverse transcription step. Only preparations negative for DNA traces were used for PCR. Samples positive for DNA contamination were treated with RNase-free DNase (100 units/µg; Roche Diagnostics, Mannheim, Germany) in 100 mM sodium acetate (pH 5.0), 5 mM MgSO$_4$ until control PCR was negative.

Primers and reverse dot-blot oligonucleotides

For PCR two different mixed oligonucleotide primer (MOP) sets, MOP-1 and MOP-2, have been designed. The primer sequences correspond to highly conserved regions present in the reverse transcriptase (RT) genes of all known human endogenous and exogenous retroviruses, as well as related animal retroviruses (Fig. 1).[14-16] MOP-1 primers preferentially amplify human and mammalian type A, B, and D reverse transcriptase sequences, whereas MOP-2 primers allow the amplification of human and mammalian type C-related RT sequences as well as RT sequences of human exogenous retroviruses such as HIV,
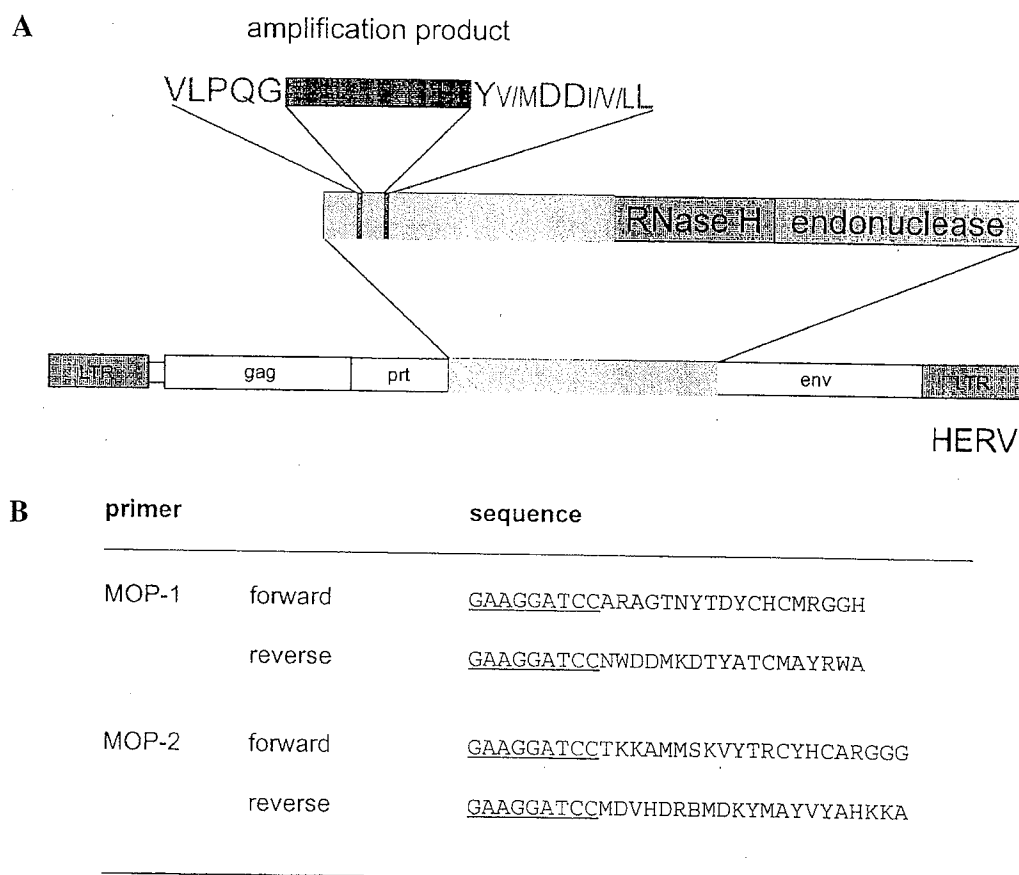

primer		sequence
MOP-1	forward	GAAGGATCCARAGTNYTDYCHCMRGGH
	reverse	GAAGGATCCNWDDMKDTYATCMAYRWA
MOP-2	forward	GAAGGATCCTKKAMMSKVYTRCYHCARGGG
	reverse	GAAGGATCCMDVHDRBMDKYMAYVYAHKKA

FIG. 1. (A) Localization of conserved amino acid domains in the amino-terminal coding region of reverse transcriptases of retroviruses. The core homology regions VLPQG and YM/V D DI/V/LL were used to design the mixed oligonucleotide primer sets MOP-1 and MOP-2. (B) Primer set MOP-1 was optimized for amplification of type A, B, and D retroviruses, whereas primer set MOP-2 was selected for favored priming of retrovirus type C-related templates as well as human exogenous retroviruses such as HIV, HTLV, HRV5, and foamy retroviruses. Standard single-letter abbreviations (IUPAC) are used to describe degenerate nucleotides. Both forward and reverse primers are oriented in the 5'-to-3' direction with respect to the DNA strand to be amplified.

HTLV, HRV5, and foamy viruses (Fig. 1). For each primer set a separate PCR was performed. The amplification products were then mixed in equimolar amounts and used as probe in the reverse dot-blot hybridization.

To design the oligonucleotides bound on the reverse dot-blot filters, databases were screened for RT-related sequences. RT sequences of exogenous and endogenous retroviruses were classified according to the current nomenclature and further subgrouped with respect to their degree of nucleotide homology (data not shown). Representative members of all retrovirus families published so far were selected (Table 1) and the sequence information corresponding to the 90-bp stretch between the highly conserved RT motifs LPQG and YM/VDDI/V/LL[16] was used for synthesis of a pair of oligonucleotides, each 45 nucleotides in length. Thus, each dot consists of an equimolar mixture of two 45-mer oligonucleotides covering the internal sequence of the MOP amplicon.

Reverse transcription and MOP PCR

Five hundred nanograms of DNA-free mRNA preparations was reverse transcribed in a volume of 50 μl containing 20 mM Tris-HCl (pH 8.4), 10 mM dithiothreitol (DTT), 50 mM KCl, 2.5 mM MgCl$_2$ deoxynucleoside triphosphates (dNTPs; 0.5 mM each), 10 units of RNasin (Promega, Madison, WI), 30 pmol of random hexamer oligonucleotides (Promega), and 20 units of murine leukemia virus (MuLV) reverse transcriptase (GIBCO-BRL, Gaithersburg, MD) at 37°C for 1 hr. Consecutively, reverse-transcribed samples were denatured for 5 min at 95°C and stored at −20°C.

MOP amplification was carried out in a total volume of 50 μl containing 1/20 of the cDNA reactions, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl$_2$, 0.001% gelatin, 50 pmol of each mixed oligonucleotide primer set, a 0.25 mM concentration of each deoxynucleoside triphosphate, and 1.25 units of Taq polymerase (GIBCO-BRL). Amplification was performed in a DNA thermal cycler (Perkin-Elmer Cetus, Emeryville, CA). Cycle parameters were as follows: 30 cycles of 94°C for 30 sec, 50°C for 4 min, and 72°C for 1 min, followed by a final extension step of 7 min at 72°C. We have chosen 50°C for the annealing step since it corresponds to the annealing temperature of the degenerate primers with the highest A–T ratio. A control reaction in which the template was omitted was carried out to detect product carryover and any traces of contaminating genomic DNA in the solutions used. Amplification products were analyzed on preparative 2.5% Tris–borate–EDTA (TBE) agarose gels and stained with ethidium bromide. Bands of interest with sizes between 100 and 150 bp corresponding to amplified retroviral reverse transcriptase sequences were excised from the gel and purified with a Gene Clean II kit (Bio 101, Vista CA). For reverse dot-blot hybridization about 50 ng of the purified fragments was labeled with [α-^{32}P]dATP (3000 Ci/mmol), using a Megaprime DNA labeling kit (Amersham Pharmacia Biotech, Little Chalfont, England).

Preparation of filter arrays

Retrovirus-specific synthetic oligonucleotides corresponding to the 90-bp internal part of the amplified RT sequence were synthesized and high-performance liquid chromatography (HPLC) purified by Birsner & Grob-Biotech GmbH (Freiburg,

Germany). For each retroviral sequence 100 pmol of a pair of 45-mer oligonucleotides mixed in equimolar amounts was diluted in 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and spotted onto ZETAprobe GT blotting membranes (Bio-Rad, Hercules, CA), using a Minifold 1 dot blotter SRC96D (Schleicher & Schuell, Dassel, Germany). Filters were rinsed in 2× SSC and oligonucleotides were irreversibly immobilized by UV cross-linking (Stratalinker; Stratagene, La Jolla CA). Filters were allowed to air dry.

Hybridizaton procedures

Standardized hybridization conditions were as follows: Prehybridization of reverse dot-blot filters was performed within sealed plastic bags in 0.25 M Na$_2$HPO$_4$ (pH 7.2), 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 50°C for at least 3 hr. For hybridization 5 × 10^5 cpm of labeled probe per milliliter of hybridization volume was added to the same solution and incubated for 16 hr under the same conditions. The membranes were then washed twice (30 min each) at 50°C in 40 mM Na$_2$HPO$_4$ (pH 7.2), 5% SDS, 1 mM EDTA and twice in 40 mM Na$_2$HPO$_4$ (pH 7.2), 1% SDS, 1 mM EDTA, respectively. Filter membranes were exposed to X-ray film (BioMax; Eastman Kodak, Rochester, NY).

RESULTS

Design of mixed oligo primers

The pol genes of all retroviruses and most retroelements share highly conserved core homology regions.[15–17] Two of the most conserved amino acid regions are the VLPQG and YV/M DDI/V/LL motifs (Fig. 1). The spacing between both motifs is about 90 base pairs and this region shows considerably less homology when compared between different retrovirus families. According to a general principle outlined by Shih et al.[15] we derived from these motifs universal PCR primers that allow amplification of all known retroviral RT-related templates. After comparison of RT core homology regions of all human and mammalian endogenous and exogenous retroviral sequences available in the database, two sets of degenerate pol primers were designed. Primer set MOP-1 was optimized for amplification of type A, B, and D retroviruses, whereas primer set MOP-2 was selected for favored priming of retrovirus type C-related templates as well as human exogenous retrovirus such as HIV, HTLV, HRV5 and foamy retroviruses. A 9-base extension featuring a clamp and a BamHI restriction site was incorporated at the 5' end of each primer. Since the sequence extension has a stabilizing effect on primer–template binding kinetics, the products generated after the first PCR cycle are amplified more efficiently in the remaining cycles. Therefore, the amplification reaction can be considered as "multiplex" PCR under moderate primer–template annealing conditions. Retroviral templates in the reaction mixture can be amplified sufficiently with MOP-1 and MOP-2 primers even when the exactly matching primer is not available. Moreover, rapid product cloning for sequence verification or characterization of novel RT-related sequences is possible. PCR conditions were optimized with respect to the amount of primers, annealing time, and annealing temperature (data not shown).

Retrovirus family	Member	Sequence	Dot code
Type B retroviruses	HERV-K(HML-1)	HML-1 (U35102)	1A
		Seq29 (S77579)	1B
	HERV-K(HML-2)	HERV-K10 (M14123)	2A
		HERV clone M3.5 (U87592)	2B
	HERV-K(HML-3)	HML-3 (U35236)	3A
		HERV1 (S66676)	3B
		RT244 (S77583)	3C
		Seq26[a]	3D
		Seq34[a]	4A
		Seq42[a]	4B
		Seq43[a]	4C
	HERV-K(HML-4)	HERV-K-T47D (AF020092)	5A
		Seq05[a]	5B
		Seq10[a]	5C
	HERV-K(HML-5)	HML-5 (U35161)	6A
	HERV-K(HML-6)	HML-6 (U60269)	7A
		Seq38[a]	7B
		Seq56[a]	7C
	HERV-K(C4)	HERV-K-C4 (U07856)	8A
		Seq31[a]	8B
	Unassigned	SeqU39937 (U39937)	5F
Type C retroviruses	HERV-H	SeqG46.2 (AF026252)	2J, 2K[b]
		Seq61[a]	2K, 5J[b]
		Seq66[a]	2L, 5K[b]
	ERV9/HERV-W	ERV9 (X57147)	4E
		Seq49[a]	4F
		Seq59[a]	4G
		Seq60[a]	4H
		Seq63[a]	4I
		Seq64[a]	4J
		HERV-W (AF009668)	4L
	ERV-FRD	ERV-FRD (U27240)	3E
		Seq46[a]	3F
	HERV-ERI	HERV-E(4-1) (M10976)	2H
		Seq32[a]	2I
	HERV-IP	HERV-I (M92067)	3H
		HERV-IP-T47D (U27241)	3I
		Seq65[a]	3J
	HERV-T	S71 pCRTK1 (U12970)	2E
		S71 pCRTK6 (U12969)	2F
Type D retroviruses	MPMV	Seq36[a]	5H
Foamy virus related		HERV-L (G895836)	1E
		Seq39[a]	1F
		Seq40[a]	1G
		Seq45[a]	1H
		Seq48[a]	1I
		Seq51[a]	1J
		Seq58[a]	1K
Unassigned human retroviral elements		Seq35[c]	5G
		Seq41[a]	5I
		Seq77[a]	5J
Human nonviral retroposon		LINE-1 (M80343)	3L
Human exogenous retroviruses		HRV5 (U46939)	6E
		Foamy virus (Y07725)	6F
		HTLV-I[c]	6G
		HTLV-II (M10060)	6H
		HIV-1[c]	6I
		HIV-2 (J04542)	6J
Mammalian endogenous retroviruses		MMTV (M15122)	7E
		PERV (AF038600)	7F
		BaEV (D10032)	7G
		GaLV (M26927)	7H
		Mo-MuLV (J02255)	7I
		MPMV (ML2349)	7J

[a]From Ref. 21.
[b]Filter code corresponding to Fig. 4 only.
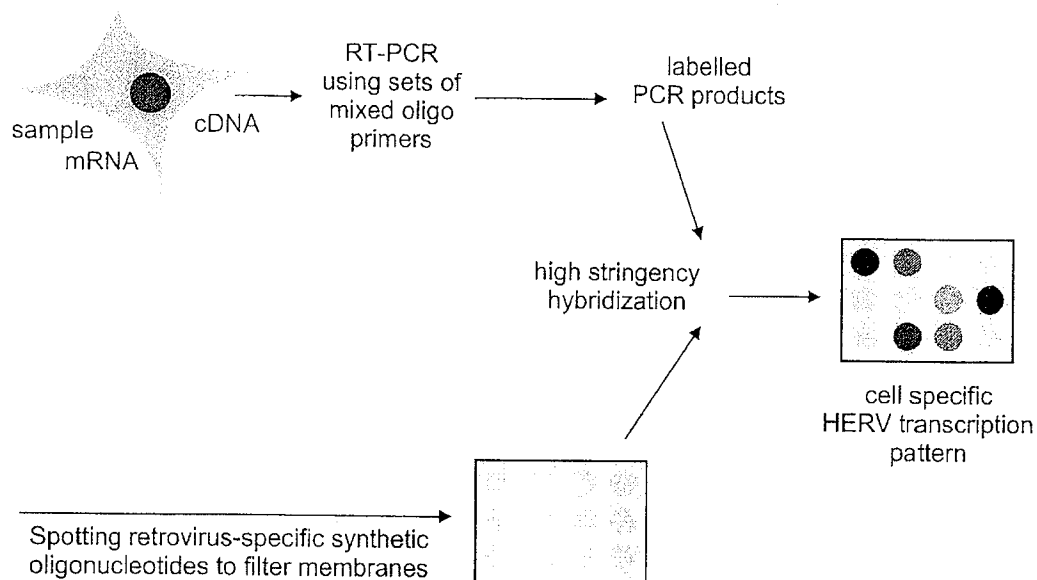[c]From Ref. 22.

FIG. 2. Flow chart of RT-PCR/RDBH procedure described in Materials and Methods. In general, samples from any biological source, e.g., all type of body fluids, tissues, cells, and cell culture supernatants, can be tested for the presence of retroviral nucleic acids.
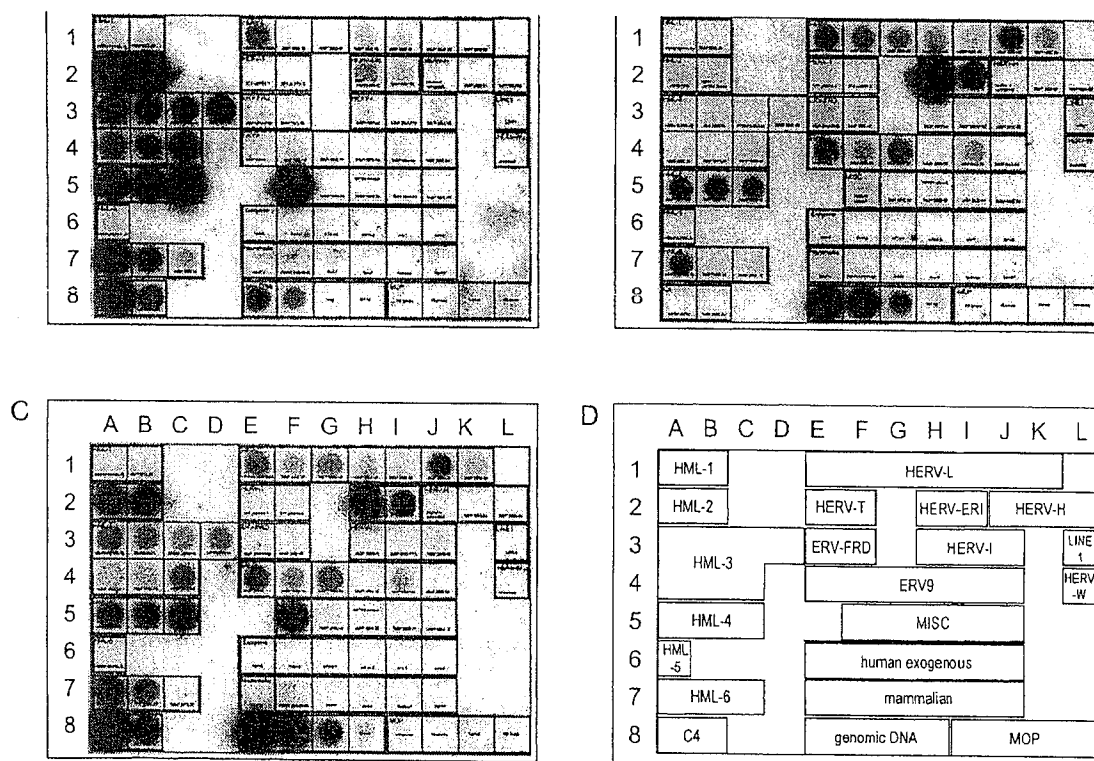

FIG. 3. HERV expression pattern in human PBMCs of a healthy blood donor. Reverse dot blots were probed under standardized conditions (as described in Materials and Methods) and DNA fragments amplified from PBMC-derived cDNA with primer set MOP-1 (**A**), MOP-2 (**B**), and with an equimolar amount of amplification products obtained with MOP-1 and MOP-2 primers, respectively (**C**). For fast assignment of retrovirus-specific oligonucleotides compare with (**D**); for origin and exact identification see Table 1.

Reverse dot-blot hybridization

In the second step, identification of the amplified products was performed by reverse dot-blot hybridization (RDBH) analysis (Fig. 2). This method is employed to sort unerringly all products of the PCR amplification. In contrast to the relaxed primer–template binding allowed during PCR amplification, RDBH enables the strict discrimination of PCR products and makes preceding false amplification of non-RT-related sequences irrelevant. The high stringency of RDBH was achieved by the use of synthetic HERV-specific oligonucleotides spotted onto the filter membranes (Table 1). These oligonucleotides correspond to the *pol* sequences amplified with MOP-1 and MOP-2 primer sets, except that they lack the primer sequences themselves. Therefore, specificity of hybridization is due solely to the amplified sequence found between the described RT core homology motifs. Thus, under high-stringency conditions the exact identification of even closely related retroviral sequences is possible.

In this study we selected 61 retrovirus-specific oligonucleotides for RDBH analysis corresponding to *pol* genes of representative members of all known human exogenous and endogenous retrovirus families. In addition, six mammalian retroviruses were included. Origin and taxonomic classification of all sequences used are summarized in Table 1.

HERV transcription pattern in human peripheral blood mononuclear cells

To assess the feasibility and specificity of the PCR/RDBH assay system, HERV transcription was analyzed in human PBMCs (Fig. 3). When the MOP-1 primer set alone was used for amplification, almost exclusively type B-related HERVs were detected, particularly members of the HERV-K subgroups HML-2, -3, -4, -6, and -C4 (Fig. 3A, dots 2A and 2B, dots 3A–3D and 4A–4C, dots 5A–5C, dots 7A and 7B, and dots 8A and 8B) and a not yet classified HERV-K-related sequence (dot 5F). This expression pattern concurs with previously published studies demonstrating a differential expression of HERV-K elements in human tissues.[18,19] No crosshybridization was observed with type C-related HERVs. Low amounts of products were obtained for one of the human foamy virus-related HERV-L elements (Fig. 3A, dot 1E).

With MOP-2 primers HERV-E-related elements (Fig. 3B; dots 2H and 2I), sequences of the HERV-L family (dots 1E–1K), and ERV9-related HERVs (dots 4E–4G, and 4I) were preferentially amplified. A certain amount of HERV-K(HML-4)- and HERV-K(HML-6)-related sequences was also present in the hybridization probe. Although the same amount of radioactively labeled probes has been used in all hybridization reactions, genomic control DNA (dots 8E–8H) gives much stronger signals with the MOP-2-amplified probe than with the MOP-1 probe, indicating that the human genome contains significantly more copies of type C-related than type B-related HERV elements.

For detection of all retroviral sequences in a single experiment MOP-1 and MOP-2 primers were first added in an equimolar ratio to the PCR. However, this experiment resulted in a predominant amplification of type C-related sequences, the ABD-type sequences being underrepresented (data not shown). Therefore, we performed separate PCR with either MOP-1 or MOP-2 primer sets, and mixed the purified amplification products of both reactions in equimolar amounts. This procedure resulted in a signal pattern that would have been expected when combining both primer sets (Fig. 3C) and corresponds roughly to the amount of type B- and type C-related HERV transcripts
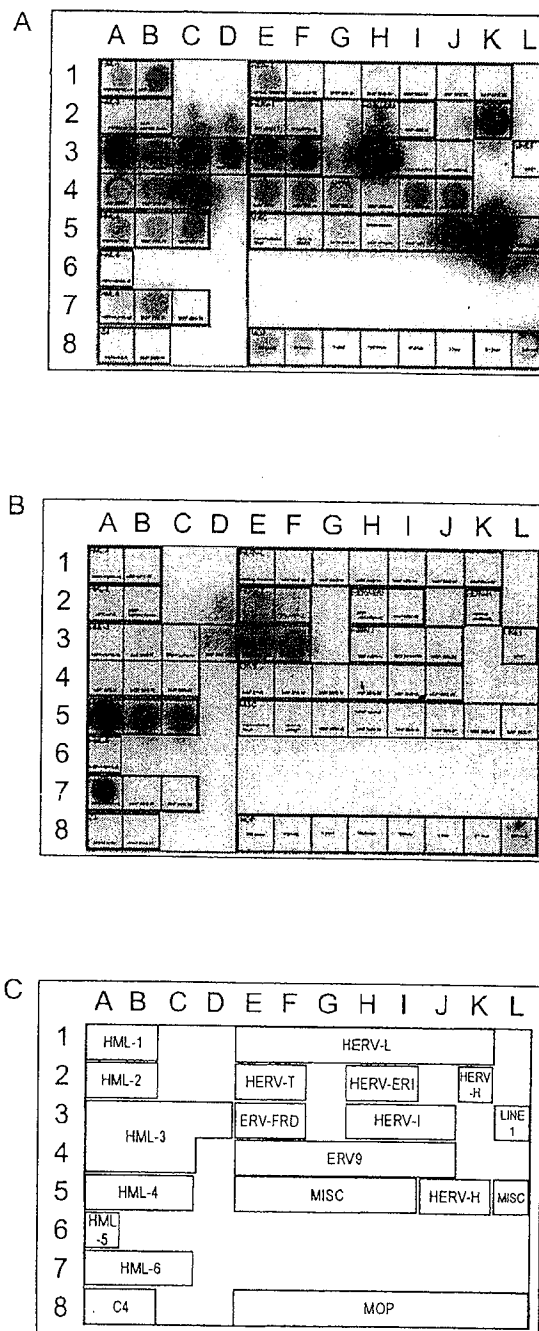

FIG. 4. HERV expression pattern in T47D cells after steroid treatment (A) and HERV transcripts packaged in T47D particles (B). DNA probes for reverse dot-blot hybridization were generated by reverse transcription of mRNA isolated from steroid-treated T47D cells (A) and from T47D particles (B), respectively.[3] For fast assignment of retrovirus-specific oligonucleotides compare with (C); for origin and exact identification see Table 1.

primer design two of the most highly conserved motifs within the reverse transcriptase-enco ding region of the *pol* gene were exploited. The usefulness of these conserved motifs for detection of novel retroviruses has been demonstrated in pioneering work by several authors.[15,16,21,22] In the first step of our experimental approach, which can be considered as "multiplex" PCR under moderate primer–template annealing conditions, all retroviral templates of the sample investigated are amplified. In the second step, exact sorting of the amplified products is performed by RDBH under high-stringency conditions. With this highly sensitive and species-specific diagnostic tool we could detect as few as 10 PERV DNA copies contained in cDNA derived from 25 ng of human PBMC mRNA.

However, it is important to emphasize that the PCR/RDBH assay is primarily a qualitative detection technique. Although the distinct intensity of the autoradiograph signals may give a strong impression of a quantitative monitoring of retrovirus expression, it is worth noting that several parameters of uncertainty may lead to a signal pattern that differs from the true expression rates in the sample. The use of highly degenerate primer sets combined with relaxed primer–template binding conditions may lead to preferential amplification of certain "high copy" or a few "best fit" templates, whereas others stay underrepresented. This effect increases with the number of PCR cycles performed and becomes critical above 35 cycles. Thus, no more than 30 rounds of PCR should be performed. The multiplex PCR does not allow an internal standardization except for overall hybridization efficacy and autoradiograph exposure time. With PCR/RDBH identified RT related transcripts must be quantified in further experiments by conventional methods such as Northern blotting or by a specific competitive PCR established for the retroviral sequence of interest.

On the other hand, it is an advantageous feature of the PCR/RDBH assay that, because of the highly degenerate primers and cross-hybridizatio n by lowering the stringency of hybridization conditions, it allows isolation and characterization of yet unknown retroviral sequences. DNA hybridizing to the covalently bound oligonucleotides can be eluted from the filter membrane by alkaline denaturation and reamplified to provide sufficient double-stranded DNA for cloning and subsequent sequence analysis.

With the employment of nonradioactive labeling techniques the PCR/RDBH assay offers the possibility of an automatable procedure for rapid analysis of retroviral expression. DNA chip technology may be applied, facilitating handling and increasing efficacy of reverse dot-blot filter membranes. Computer-assisted evaluation of RDBH results by phospho/fluorescence-imaging systems may further improve visualization. It is one of the advantageous features of the PCR/RDBH assay system that the test is unlimited with respect to number and origin of retroviral RT sequences to be tested. Novel RT-encoding sequences can be easily added to the filter arrays. Modifications in the experimental design are not necessary. Moreover, by modifying the hybridization conditions PCR/RDBH can be used to search for new exo- or endogenous retroviruses with only weak homologies to already known families.

In summary, the PCR/RDBH assay is a powerful technique for precise qualitative analysis of retrovirus activity in biological samples. Number and types of retroviral sequences to be identified are determined solely by number and types of synthetic oligonu-

cleotides spotted onto reverse dot-blot membranes. PCR/RDBH could be useful in guarding patients against undesired transmission of genetic material by retroviruses from therapeutic protein preparations, in gene therapy and xenotransplantation.

ACKNOWLEDGMENTS

REFERENCES

1. Wilkinson DA, Mager DL, and Leong JC: Endogenous human retroviruses. In: The Retroviridae (Levy JA, ed.), Vol. 3. Plenum Press, New York, 1994, pp. 465–535.
2. Leib-Mösch C and Seifarth W: Evolution and biological significance of human retroelements. Virus Genes 1996;11:133–145.
3. Seifarth W, Skladny H, Krieg-Schneider F, Reichert A, Hehlmann R, and Leib-Mösch C: Retrovirus-like particles released from the human breast cancer cell line T47-D display B- and C-type related endogenous retroviral sequences. J Virol 1995;69:6408–6416.
4. Seifarth W, Baust C, Murr A, Skladny H, Krieg-Schneider F, Blusch J, Werner T, Hehlmann R, and Leib-Mösch C: Proviral structure, chromosaomal location and expression of HERV-K-T47D: A novel human endogenous retrovirus derived from T47D particles. J Virol 1998;72:8384–8391.
5. Patience C, Takeuchi Y, and Weiss RA: Infection of human cells by an endogenous retrovirus of pigs. Nature Med 1997;3:282–286.
6. Patience C, Takeuchi Y, Cosset FL, and Weiss RA: Packaging of endogenous retroviral sequences in retroviral vectors produced by murine and human packaging cells. J Virol 1998;72:2671–2676.
7. Günzburg WH and Salmons B: Development of retroviral vectors as safe, targeted gene delivery systems. J Mol Med 1996;74:171–182.
8. Martin U, Kiessig V, Blusch JH, Haverich A, van der Helm K, Herden T, and Steinhoff G: Expression of pig endogenous retrovirus by primary porcine endothelial cells and infection of human cells. Lancet 1998;352:692–694.
9. Wilson CA, Wong S, Muller J, Davidson CE, Rose TM, and Burd P: Type C retrovirus released from porcine primary peripheral blood mononuclear cells infects human cells. J Virol 1998;72:3082–3087.
10. Takeuchi Y, Patience C, Magre S, Weiss RA, Banerjee PT, Le Tissier P, and Stoye JP: Host range and interference studies of three classes of pig endogenous retrovirus. J Virol 1998;72:9986–9991.
11. Bach FH and Fineberg HV: Call for a moratorium on xenotransplants. Nature (London) 1998;391:326.
12. Stoye JP and Coffin JM: The dangers of xenotransplatatio n. Nature Med 1995;1:1100.
13. Sambrook J, Fritsch EF, and Maniatis T: *Molecular Cloning: A Laboratory Manual,* 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
14. Xiong Y and Eickbush TH: Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 1990;9:3353–3362.

15. Shih A, Misra R, and Rush MG: Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: Relation to primate retroviruses. J Virol 1989;62:64–75.

16 Donehower LA, Bohannon RC, Ford RJ, and Gibbs RA: The use of primers from highly conserved *pol* regions to identify uncharacterized retroviruses by the polymerase chain reaction. J Virol Methods 1990;28:33–46.

17. McClure MA, In Skalka AM, and Goff SP (eds.): *Evolutionary History of Reverse Transcriptase.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1993: pp. 425–444.

18. Mestrand P, Lindeskog M, and Blomberg J: Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. J Gen Virol 1992;73:2463–2466.

19. Andersson ML, Medstrand P, Yin H, and Blomberg J: Differential expression of human endogenous retroviral sequences similar to mouse mammary tumor virus in normal peripheral blood mononuclear cells. AIDS Res Hum Retroviruses 1996;12:833–840.

20. Krieg AM, Gourley MF, Klinman DM, Perl A, and Steinberg AD: Heterogenous expression and coordinate regulation of endogenous retroviral sequences in human peripheral blood mononuclear cells. AIDS Res Hum Retroviruses 1992;8:1991–1998.

21. Herrmann M, and Kalden JR: PCR and reverse dot blot hybridization for detection of endogenous retroviral transcripts. J Virol Methods 1994;46:333–348.

22. Tuke PW, Perron H, Bedin F, Beseme F, and Garson JA: Development of a pan-retrovirus detection system for multiple sclerosis studies. Acta Neurol Scand 1997;169(Suppl.):16–21.

Address reprint requests to:
Wolfgang Seifarth
III. Medizinische Klinik
Klinikum Mannheim der Universität Heidelberg
Wiesbadener Str. 7-11
D-68305 Mannheim, Germany

E-mail: seifarth@rumms.uni-mannheim.de